Evolutionary Genomics of Resistance Gene Loci in Wheat and Barley

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Beatrice Scherrer

von

Willisau LU

Promotionskomitee

Prof. Dr. Beat Keller (Vorsitz)

Dr. Catherine Feuillet (Leitung der Dissertation)

Prof. Dr. Bruce McDonald

Zürich 2006

I. SUMMARY **1**

II. ZUSAMMENFASSUNG **4**

III. GENERAL INTRODUCTION **8**

1. Comparative genomics in grasses **8**

1.1 The grass family: origin and genome organization 8
1.2 Colinearity among grass genomes 13
1.3 Interspecific sequence comparison 14
1.4 Intraspecific sequence comparison 17

2. Evolution of plant disease resistance (*R*) gene loci **21**

2.1 Race-specific resistance in plant-pathogen interactions 21
2.2 The different classes of *R* genes 23
2.3 Organization of *R* gene loci and their evolution 24

3. Aim of the study **28**

IV. ANCIENT HAPLOTYPES RESULTING FROM EXTENSIVE MOLECULAR REARRANGEMENTS IN THE WHEAT A GENOME HAVE BEEN MAINTAINED IN SPECIES OF THREE DIFFERENT PLOIDY LEVELS **29**

1. Abstract **29**

2. Introduction **31**

3. Materials and Methods **34**

3.1 Genomic DNA isolation and PCR 34
3.2 BAC clone isolation and sequencing 34
3.3 Sequence analysis 36

4. Results **38**

4.1 Contig establishment at the *Lr10* locus in tetraploid and hexaploidwheat 38
4.2 Sequence organisation at the *Lr10* locus in tetraploid and hexaploid wheat 40
4.3 A gene-rich region is highly conserved in the three homoeologous A genomes 43
4.4 The large intergenic regions between *RGA2* and *Lr10* evolved differentially in *T. monococcum* and *T. turgidum* of the H1 haplotype 45
4.5 The two haplotypes H1 and H2 originate from ancient and extensive rearrangements 50
4.6 Identical H2 haplotypes are found at three ploidy levels and H2 sub-haplotypes result from different types of rearrangements 52

5. Disussion **57**

5.1 A large deletion/inversion event is at the origin of the H2 haplotype 58
5.2 Common themes of evolution in both haplotypes 59
5.3 Old and stable haplotypes at the *Lr10* resistance locus 61

V. LARGE INTRASPECIFIC HAPLOTYPE VARIABILITY AT THE *RPH7* LOCUS RESULTS FROM RAPID AND RECENT DIVERGENCE IN THE BARLEY GENOME 63

1. Abstract 63

2. Introduction 64

3. Materials and Methods 67

3.1 Plant material 67
3.2 Shotgun sequencing and sequence analysis 67
3.3 Identification and mapping of sorghum BACs containing genes homologous to the *Rph7* locus 68
3.4 Dating of retrotransposon insertions 69
3.5 Haplotype analysis 70

4. Results 71

4.1 Establishment of a 350-kb Contig at the *Rph7* Locus in Cepada Capa 71
4.2 Comparison of homologous contigs in Morex and Cebada Capa reveals dramatic and recent genome rearrangements at the *Rph7* locus 73
4.3 High variability is found in intergenic regions at the *Rph7* locus in the cultivated barley gene pool 80
4.4 Evolution of the gene composition at *Rph7* orthologous loci in wheat, barley, rice, and sorghum 85

5. Discussion 89

5.1 Intraspecific comparison of homologous loci reveals mechanisms underlying rapid genome evolution in barley 89
5.2 Large haplotype variability in the cultivated barley gene pool 93
5.3 A complex history of rearrangements involving gene movements is responsible for the gene diversity at the *Rph7* locus in the Triticeae 94

VI. FUNCTIONAL ANALYSIS OF THE CANDIDATE GENES FOR *RPH7* 96

1. Introduction 96

2. Materials and Methods 100

2.1 Plasmid constructs for *Agrobacterium*-mediated barley transformation 100
2.2 *Agrobacterium*-mediated barley transformation 101
2.3 Characterization of transgenic plants 101
2.4 Analysis of gene expression in transgenic plants 102
2.5 Virus-induced gene silencing in barley 103
2.6 Screening for leaf rust susceptible mutant plants 105
2.7 Re-evaluation of recombination events with genomic STS markers 105

3. Results and Discussion **107**

3.1 *Agrobacterium*-mediated transformation of the susceptible barley cv. Golden Promise 107
3.2 Expression analysis of the transgenes by RT-PCR 110
3.3 Virus-induced gene silencing of the two candidate genes *Hvpg1* and *HvHGA1* 112
3.4 Identification of γ-irradiation *Rph7* mutants 114
3.5 Possible reasons for the lack of complementation with the four candidate genes and perspectives 116

VII. GENERAL DISCUSSION **120**

1. Comparative genomics in wheat and barley as a tool to reveal mechanisms of genome evolution in the Triticeae **120**
2. A helitron transposable element is present at the *Rph7* locus in Morex **124**
3. Future prospects **127**

VIII. REFERENCES **129**
IX. APPENDIX **138**
X. ACKNOWLEDGEMENTS **143**
XI. CURRICULUM VITAE **144**

Die einzige Möglichkeit, nie zu scheitern ist,
nichts zu versuchen.

Bertrand Piccard

I. SUMMARY

Grasses (Poaceae) are a large and diverse family of monocotyledonous flowering plants that include the important staple cereals wheat, barley, rice, maize, sorghum, millet, and oat. Beyond their agronomic importance, grasses also serve as a model system for comparative genetics and genomics. Despite tremendous variations in genome size, ploidy level, and chromosomal number, comparative mapping studies have demonstrated that the linear order of genes (colinearity) is largely conserved between different grass species. However, comparing orthologous genomic segments at the DNA sequence level (microcolinearity) has revealed many exceptions to colinearity due to local rearrangements between otherwise colinear regions. These microcolinearity studies allowed the characterization of some of the molecular mechanisms that have shaped the grass genomes during evolution. Surprisingly, intraspecific sequence comparisons have recently revealed that genomic rearrangements do not only occur between different grass species but also between different lines or varieties within the same species.

The aim of this thesis was to analyze the genomic organization and the molecular evolution of two leaf rust disease resistance loci *Lr10* and *Rph7* in wheat and barley, respectively. In the first part of the study, a detailed sequence comparison of the *Lr10* locus in the A genomes of diploid (*T. monococcum*), tetraploid (*T. turgidum* ssp *durum*), and hexaploid (*T. aestivum*) wheat was performed. Two haplotypes, H1 and H2, defined by the presence or absence of the two genes *Lr10* and *RGA2* at this locus, were compared at the sequence level. Conservation between the two haplotypes was limited to the genes located on both sides of a large 150 kb region including the *Lr10* and *RGA2* genes. Within this region,

extensive rearrangements such as retroelement insertions, transposon deletions and unequal recombination within elements occurred. A large deletion/inversion event that eliminated the disease resistance gene *Lr10* and a part of *RGA2* is assigned to the origin of the H2 haplotype. Gene conservation between the H1 haplotype of diploid and tetraploid wheat and the conservation of the rearrangements that resulted in the H2 haplotype at three ploidy levels indicated that the two haplotypes are of ancient origin and were exceptionally stable during wheat evolution.

In the second part of the work, a physical BAC contig of 350 kb was established and sequenced at the *Rph7* locus in the leaf rust resistant barley cultivar Cebada Capa. The sequence was compared to a previously sequenced homologous BAC contig from the susceptible cultivar Morex. Less than 35% of the two sequences were conserved and colinearity was restricted to five genic and two intergenic regions. The intervals separating the seven conserved regions varied in size and completely different numbers and types of repetitive elements were found in the homologous Cebada Capa and Morex sequences. In both cultivars, the non-conserved regions consisted of about 53% repetitive sequences, mainly composed of long-terminal repeat (LTR) retrotransposons that have inserted less than 1 million year ago. A PCR-based analysis of intergenic regions at the *Rph7* locus and three additional loci in 41 barley cultivars revealed large haplotype diversity in the cultivated barley gene pool and indicated that homologous loci diverged recently and rapidly in barley.

The sequence comparison between Morex and Cebada Capa BAC contigs also aimed at identifying possible additional candidate genes for *Rph7* in the resistant cultivar compared to those previously found in the susceptible sequence of Morex. No other gene was detected on the Cebada Capa sequence indicating that the four genes *Hvpg1, Hvpg4, HvHGA1,* and *HvHGA2* were candidates for the resistance

gene. In the third part of the thesis, the four candidate genes were functionally analyzed by *Agrobacterium tumefaciens*-mediated transformation of the susceptible barley cultivar Golden Promise. For each of the four genes about ten T_1 families derived from individual T_0 plants with a single insertion of the transgene were pheno- and genotypically characterized. None of the 520 transgenic T_1 plants that were analyzed showed a resistance phenotype, indicating that none of the four candidate genes was able to complement the susceptible phenotype of cv. Golden Promise. RT-PCR analysis demonstrated that the transgenes are expressed in the T_1 plants and thereby that the absence of resistance reaction after leaf rust infection was not due to a lack of transgene expression. A virus induced gene silencing (VIGS) experiment set up for *Hvpg1* and *HvHGA1* confirmed the absence of complementation observed in the stable transformation experiment. Re-analysis of the recombination breakpoints delimiting the genetic interval comprising *Rph7* with new genomic markers confirmed previous mapping data and the presence of the resistance gene in this genetic interval. Thus, further investigations will be needed to identify *Rph7* from the sequence of Cebada Capa.

II. ZUSAMMENFASSUNG

Gräser (Poaceae) bilden eine grosse und vielfältige Familie von einkeimblättrigen Blütenpflanzen, welche die agronomisch wichtigen Getreide Weizen, Gerste, Reis, Mais, Hirse und Hafer beinhaltet. Gräser sind nicht nur in agronomischer Hinsicht von Bedeutung, sondern dienen auch als Modellsystem für vergleichende Genetik und Genomik. Vergleichende Kartierungsstudien haben gezeigt, dass die lineare Anordnung der Gene (Kolinearität) in Gräsern trotz unterschiedlicher Genomgrössen, Ploidiegrade und Anzahl der Chromosomen weitgehend konserviert ist. Der Vergleich orthologer Abschnitte auf Ebene der DNA-Sequenz (Mikrokolinearität) hat jedoch viele Ausnahmen zur allgemeinen Kolinearität aufgrund von lokalen Umgestaltungen zwischen ansonsten kolinearen Regionen aufgedeckt. Diese Mikrokolinearitätsstudien ermöglichten es, einige der molekularen Mechanismen, welche Gräsergenome während der Evolution geformt haben, zu charakterisieren. Sequenzvergleiche innerhalb einer Art haben kürzlich gezeigt, dass genomische Reorganisationen überraschenderweise nicht nur zwischen verschiedenen Grasspezies, sondern auch zwischen verschiedenen Linien oder Varietäten einer Art auftreten.

Das Ziel der vorliegenden Arbeit war, die genomische Organisation und die molekulare Evolution der beiden Braunrost-Resistenzloci *Lr10* in Weizen und *Rph7* in Gerste zu analysieren. Im ersten Teil der Studie wurde ein detaillierter Sequenzvergleich des *Lr10* Locus im A Genom von diploidem (*T. monococcum*), tetraploidem (*T. turgidum* ssp *durum*) und hexaploidem (*T. aestivum*) Weizen durchgeführt. Zwei Haplotypen, H1 und H2, welche durch die Präsenz oder das Fehlen der zwei Gene *Lr10* und *RGA2* an diesem Locus definiert sind, wurden auf

Sequenzebene verglichen. Die Konservierung zwischen den beiden Haplotypen beschränkte sich auf Gene beidseitig einer 150 kb langen Region, in welcher die Gene *Lr10* und *RGA2* liegen. Innerhalb dieser Region haben umfangreiche Umstrukturierungen wie beispielsweise Insertionen von Retroelementen, Deletionen von Transposons und ungleiche Rekombination innerhalb von Elementen stattgefunden. Ein grosses Deletions-/Inversionsereignis, wodurch das Krankheitsresistenzgen *Lr10* und ein Teil von *RGA2* eliminierte wurden, liegt dem Ursprung des H2 Haplotypen zu Grunde. Sowohl die Genkonservierung zwischen den H1 Haplotypen von diploidem und tetraploidem Weizen als auch die Konservierung der Umordnungen, welche zum H2 Haplotypen geführt haben, deuten darauf hin, dass die beiden Haplotypen sehr alt sind und während der Weizenevolution konserviert wurden.

Im zweiten Teil der Studie wurde ein 350 kb langes physikalisches BAC Kontig am *Rph7* Locus im Braunrost-resistenten Gerstenkultivar Cebada Capa erstellt und sequenziert. Diese Sequenz wurde mit einem früher sequenzierten homologen BAC Kontig aus dem anfälligen Kultivar Morex verglichen. Weniger als 35% der beiden Sequenzen waren konserviert und die Kolinearität beschränkte sich auf fünf genische und zwei inter-genische Regionen. Die Intervalle, welche die sieben konservierten Regionen voneinander trennen, hatten unterschiedliche Längen und sowohl die Anzahl als auch die Art der gefundenen repetitiven Elemente war völlig verschieden zwischen den homologen Sequenzen von Cebada Capa und Morex. In beiden Kultivaren bestehen die nicht-konservierten Regionen aus rund 53% repetitiven Sequenzen, vorwiegend aus Long-terminal Repeat (LTR) Retrotransposons, welche sich vor weniger als 1 Million Jahren eingefügt hatten. Eine PCR-basierte Analyse der intergenischen Regionen am *Rph7* Locus und an drei zusätzlichen Loci in 41

Gerstenkultivaren zeigte, dass im kultivierten Gerstengenpool eine grosse Haplotypdiversität besteht. Dies deutet darauf hin, dass homologe Loci in Gerste rasch und vor kurzem divergierten.

Im Weiteren sollte der Sequenzvergleich zwischen den Morex und Cebada Capa BAC Kontigs dazu dienen, neben den bereits in Morex gefunden Genen zusätzliche Kandidatengene für *Rph7* zu identifizieren, falls solche im resistenten Kultivar Cebada Capa vorhanden sein sollten. Da jedoch in der Cebada Capa Sequenz keine weiteren Gene entdeckt wurden, blieben nur die vier Gene *Hvpg1, Hvpg4, HvHGA1* und *HvHGA2* Kandidaten für das Resistenzgen. Im dritten Teil dieser Dissertation wurden die vier Kandidatengene mittels *Agrobacterium tumefaciens*-vermittelter Transformation des anfälligen Gerstenkultivars “Golden Promise“ funktionell überprüft. Für jedes der vier Gene wurden rund zehn T_1 Familien, welche von individuellen T_0 Pflanzen mit einer einzelnen Insertion des Transgens abstammten, phäno- und genotypisch charakterisiert. Keine der 520 analysierten transgenen T_1 Pflanzen zeigte einen resistenten Phänotyp. Daraus liess sich schliessen, dass keines der vier Kandidatengene fähig war, den anfälligen Phänotyp von Golden Promise zu komplementieren. Mittels RT-PCR Analyse wurde gezeigt, dass die Transgene in den T_1 Pflanzen exprimiert werden. Folglich war die ausbleibende Resistenzreaktion nach der Infektion mit Braunrost nicht auf fehlende Expression der Transgene zurückzuführen. Virus-induziertes Ausschalten (VIGS) der Gene *Hvpg1* und *HvHGA1* bestätigte die negativen Komplementierungsresultate der stabilen Transformation. Die Rekombinationsstellen, welche das genetische Intervall um *Rph7* eingrenzen, wurden daraufhin nochmals mit neuen genetischen Markern überprüft. Die früheren Kartierungsdaten und somit das Vorhandensein des Resistenzgens in diesem genetischen Intervall konnte bestätigt werden. Zur

Identifikation von *Rph7* in der Cebada Capa Sequenz werden folglich weitere Untersuchungen notwenig sein.

III. GENERAL INTRODUCTION

1. Comparative genomics in grasses

1.1 The grass family: origin and genome organization

The grass family (Poaceae) comprises around 10,000 species which are found in many ecosystems, from temperate to tropical habitats. Altogether, grasses cover more than 20% of the earth's land area. (Shantz, 1954). Not surprisingly, grasses, and in particular the cereals, are a major food source for humans and, directly and indirectly, account for ~60% of the world's food production [FAOSTAT home page; http://faostat.fao.org/]. During the last 10,000 years, humans have domesticated several grass species as crop plants, such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), rice (*Oryza sativa*), sorghum (*Sorghum bicolor*) and maize (*Zea mays*).

It is estimated that the grass family originated roughly 77 million years ago (mya) (Figure 3-1; Gaut, 2002). The angiosperm (flowering plants) lineage is thought to be ~200 million years old. Thus, the grasses represent a relatively recent branch of the plant family tree. The Poaceae are further divided into six or seven major subfamilies whereof four (Pooideae, Erhartoideae, Chloridoideae and Panicoideae) include most of the cereals. The two subfamilies Erhartoideae (rice) and Pooideae (oats, barley and wheat) are estimated to have diverged 46 mya (Figure 3-1). Within the Pooideae, different tribes containing species of agronomic interest have been classified. The Triticeae tribe (represented by wheat and barley) diverged from the Avena tribe 25 mya. Within the Triticeae, divergence time between wheat and barley is estimated to be about 13 mya (Figure 3-1).

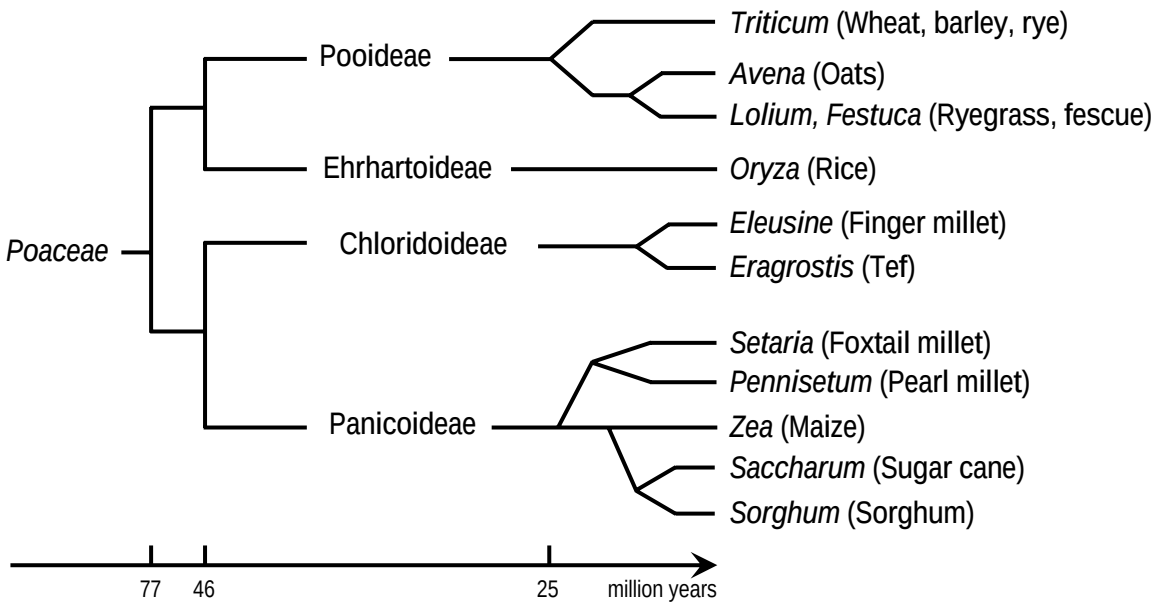


Figure 3-1 Grass phylogeny showing the four subfamilies that include most of the cereals.

The evolution process from a common ancestor to present species is illustrated. (Modified from Devos, 2005) Divergence time estimates are according to Gaut, 2002.

With regard to genome organization, grasses represent a highly diverse family, displaying extensive variation in chromosomal number, ploidy level, and genome size. Rice is diploid with 12 chromosomes (2*n* = 24) and a genome size of 420 Mbp, which is the smallest genome among the major cereal crops. Its small genome which is only approximately three times larger than the one of the dicot model plant *Arabidopsis thaliana* has established rice as the model plant for monocotyledons. Other grass species such as maize (2*n* = 20 / 2,500 Mbp), barley (2*n* = 14 / 4,900 Mbp), and hexaploid wheat (2n = 42 / 16,000 Mbp) have much larger genome sizes (Bennett and Leitch, 1995). Part of the variation in nuclear DNA content is a consequence of polyploidy (genome doubling) and it is estimated that roughly 44% of the species in the grass family are polyploids (DeWet, 1986). Moreover, recent genomic studies have revealed that most or all modern flowering plant genomes

harbor evidence of multiple rounds of past polyploidization events, often followed by massive silencing and elimination of duplicated genes (diploidization) (Adams and Wendel, 2005). Arabidopsis was considered a diploid until whole-genome sequence analysis revealed three ancestral rounds of duplications that have led to the current genome structure (Vision *et al.*, 2000; Simillion *et al.*, 2002; Blanc *et al.*, 2003). A similar analysis of the rice genomic sequence has recently demonstrated that at least 53-62% of the sequence, and possibly the entire rice genome, results from an ancient duplication (Vandepoele *et al.*, 2003; Guyot and Keller, 2004; Wang *et al.*, 2005). Phylogenetic dating of duplicated rice genes suggested that the duplication event occurred about 70 million years ago, *i.e.* it predates the divergence of the major grass lineages (Paterson *et al.*, 2004). This indicates that all cereal crops, even those considered to be diploids nowadays, are likely of polyploid origin. Polyploidy has been proposed as a major force of genome evolution because it provides the genetic raw material for DNA diversification and allows species to evolve new functions and adapt to a wider range of habitats and environmental conditions (Ohno, 1970). Blanc and Wolfe (2004) have recently demonstrated that some classes of genes (such as those involved in transcription and signal transduction) have been preferentially retained after the most recent polyploidization event in Arabidopsis whereas other classes (including those involved in DNA repair and those for organellar proteins) have been preferentially lost. Moreover, functional analysis of recently duplicated gene pairs revealed that more than half of the pairs formed by polyploidy have significantly different expression patterns and 62% of the gene pairs have undergone functional diversification (Blanc and Wolfe, 2004). In addition to naturally occurring polyploids, much has been learnt from newly created, synthetic polyploids. This approach has been successful for studying early consequences of

polyploidy. Extensive and rapid genomic changes, including sequence rearrangements, homoeologous recombination, sequence elimination, and changes in DNA methylation have been reported in several systems. Some of these changes have occured immediately after polyploidization, whereas others have occurred within a few generations (reviewed in Adams and Wendel, 2005). These studies also indicated that gene silencing is a common response to polyploidization although it is not yet clear how and why duplicated genes are silenced in polyploids.

More important for the variability in genome size among grass species than polyploidy, however, are the differences in the amount of repetitive DNA which can account for up to 80% of the DNA in large genome species such as barley and wheat. Most of the repetitive DNA consists of retroelements, primarily nested long-terminal repeat (LTR)-retrotransposons (SanMiguel *et al.*, 1996; Bennetzen, 2002). Because of their copy-and-paste mode of transposition, these elements tend to accumulate while they are active and thereby increase genome size (Figure 3-2A). The first evidence for the role of LTR-retrotransposons in genome size difference was reported after the sequencing and comparison of the region around the alcohol dehydrogenase *adh1* gene in maize and sorghum, demonstrating that the two species vary threefold in length at this locus (Tikhonov *et al.*, 1999). Retrotransposons accounted for more than 74% of the maize *adh1* region, whereas the corresponding region in sorghum did not contain any retroelements. Comparative analysis and dating of insertion time have shown that the insertion of transposable elements occurred recently in terms of evolutionary time-scales. Indeed, more than 80% of the intact LTR-retrotransposons analyzed in angiosperms have been inserted in the past five million years (Bennetzen, 2005). The time of insertion of a LTR-retrotransposon can be estimated by comparing the two LTR sequences. When a

particular retrotransposon inserts into genomic DNA, both copies of the LTR are identical. In the course of time, the LTRs accumulate nucleotide substitutions and consequently diverge in sequence. Assuming that the accumulation of nucleotide substitutions occurs at a regular pace, the number of nucleotide differences between the two LTRs allows to determine the time span that has passed since the element inserted (SanMiguel *et al.*, 1998).

Genome size increase is counteracted by deletion mechanisms that reduce genome size. For example, unequal recombination between the LTRs of a single retroelement or between the LTRs of different elements can remove the internal domain and result in a solo-LTR (Figure 3-2B). In barley, a 16-fold excess of LTRs relative to internal retroelement domains was observed for the BARE-1 element, indicating that conversion to solo-LTRs occurs more rapidly than integration of new LTR-retrotransposons (Vicient *et al.*, 1999; Shirasu *et al.*, 2000). In addition to the intra- and inter-element crossover which results in solo-LTRs, deletion of large random fragments of repetitive DNA was reported in wheat, rice and Arabidopsis (Wicker *et al.*, 2001; Devos *et al.*, 2002; Ma *et al.*, 2004). These deletions were ascribed to illegitimate recombination, a phenomenon that can act anywhere because it requires as little as 1 bp of sequence homology (Bennetzen *et al.*, 2005). Analysis of the Arabidopsis genome for evidence of genomic DNA loss indicated that illegitimate recombination is the driving force behind genome size decrease in Arabidopsis, removing at least fivefold more DNA than unequal homologous recombination (Devos *et al.*, 2002).

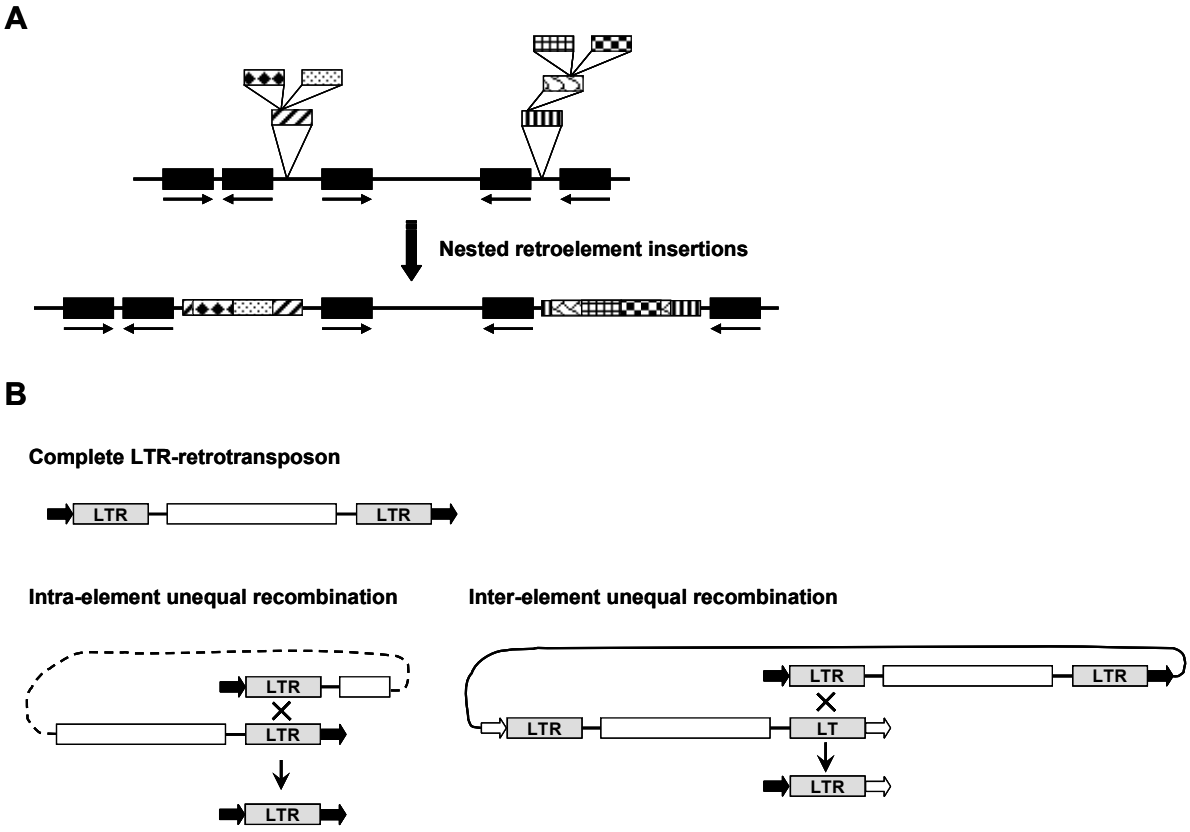


Figure 3-2 Possible mechanisms of genome expansion and contraction.

(A) Insertion of retroelements in intergenic regions is a major driving force of genome expansion. The invasion can occur in several waves, leading to insertion of retrotransposons within each other (nested retroelements). Genes are represented by black boxes. Retrotransposons are indicated with patterned boxes. (B) Solo-LTR resulting from intra- (left) or inter-element (right) unequal recombination. The dotted line indicates the folding needed to accomplish the intra-element recombination and does not represent DNA whereas the solid black line represents sequences between the two LTR-retrotransposons that are deleted via the inter-element recombination.

1.2 Colinearity among grass genomes

Despite the differences in ploidy level, chromosome number, and haploid DNA content, the first comparative mapping studies that were performed in the 1990s indicated that the linear order (colinearity) of genes remained largely conserved between related plant species over millions of years of evolution, but colinearity has

been particularly well studied in grass species (reviewed in Devos and Gale, 2000; Keller and Feuillet, 2000; Feuillet and Keller, 2002; Devos, 2005). Initial comparisons in grasses, using restriction fragment length polymorphisms (RFLP) markers, revealed that the three diploid genomes (A, B, and D) that form present-day hexaploid bread wheat had retained almost identical marker and gene orders (Chao *et al.*, 1988). During the following years, close relationships were demonstrated between the genomes of almost all important grass species. In 1995, the first consensus grass map aligning the genomes of seven different grass species in terms of "rice linkage blocks" was published by Moore *et al.* (1995). Since then, the so-called "Crop Circle" has constantly been refined and enhanced with new species (for current version see Devos, 2005).

1.3 Interspecific sequence comparison

During the last decade, comparative genomics in grasses has progressed from the comparison of the marker order on genetic maps (colinearity) to more detailed assessments of the order of genes at the DNA sequence level (microcolinearity). The development of large insert libraries, such as bacterial artificial chromosome (BAC) libraries, has allowed the isolation and sequencing of large genomic fragments (100 – 500 kb) of orthologous regions from different grass genomes. One of the first microcolinearity studies was performed at the *sh2/a1* orthologous regions in maize, sorghum and rice (Chen *et al.*, 1997; Chen *et al.*, 1998). Four genes were found conserved in order and orientation, demonstrating microscale colinearity at orthologous loci of these three distantly related species. In wheat, however, the *sh2/a1* region was not colinear and a translocation had separated the *sh2* and *a1*

homologues onto separate wheat chromosomes (Li and Gill, 2002). In contrast to the *sh2/a1* locus, the region surrounding the *adh1* gene of maize, rice and sorghum has undergone substantial rearrangement (Tikhonov *et al.*, 1999). Nine genes were found in colinear order between maize and sorghum whereas three were missing from this region in maize compared to sorghum. The rice region that contains the orthologue of the maize *adh1* gene does not otherwise show any collinearity with the maize or sorghum *adh1* regions, suggesting that the *adh1* gene had moved to its current location in a common ancestor of sorghum and maize as a single gene translocation (Tarchini *et al.*, 2000; Bennetzen and Ramakrishna, 2002; Ilic *et al.*, 2003). Several further microcolinearity studies compared different orthologous loci harboring genes involved in resistance (*e.g. Rp1, Rph7*), development (*e.g. Vrn1, lg2/lrs1*), and quality (*e.g. Zein, Ha*) (Table 3-1).

Table 3-1 Gene loci for which interspecific sequence comparisons were performed.

Locus	Compared plant species	Reference
Lrk	Wheat, barley, maize, rice	(Feuillet and Keller, 1999)
Rp1	Maize, sorghum	(Ramakrishna *et al.*, 2002)
Rph7	Barley, rice	(Brunner *et al.*, 2003)
adh1/adh2	Maize, sorghum, rice	(Tikhonov *et al.*, 1999) (Tarchini *et al.*, 2000)
Vrn1	Wheat, barley, sorghum, rice	(Dubcovsky *et al.*, 2001) (Ramakrishna *et al.*, 2002)
lg2/lrs1	Maize, rice	(Langham *et al.*, 2004)
sh2/a1	Maize, sorghum, rice, wheat	(Chen *et al.*, 1997) (Chen *et al.*, 1998)
Zein gene cluster	Maize, sorghum, rice	(Song *et al.*, 2002)
Ha	Barley, rice	(Caldwell *et al.*, 2004)
r/b	Maize, sorghum, rice	(Swigonova *et al.*, 2005)

These studies demonstrated that colinearity is generally maintained at the genetic map level between orthologous genomic segments of grass genomes, though to different degrees depending on the chromosomal region and the species. However, in general, microcolinearity studies have revealed a much higher degree of diversity at the microstructural level than was predicted by comparative mapping (estimated range from 50% to 73%; Gaut, 2002). Many small-scale genic rearrangements, such as single or multiple gene insertions and/or deletions, tandem duplications, inversions, and translocations were reported that were previously overlooked by comparative mapping (reviewed in Feuillet and Keller, 2002; Gaut, 2002; Bennetzen and Ma, 2003; Guyot *et al.*, 2004). Furthermore, microcolinearity studies do not only provide valuable insight into the genome organization but also help to identify the molecular mechanisms of the rearrangements between the grass genomes and thus to understand the evolution of grass genomes. Retroelements have been reported to be involved in genome expansion by nested insertions in maize, barley, and wheat (SanMiguel *et al.*, 1996; Wicker *et al.*, 2001; Rostoks *et al.*, 2002). Different studies have shown the ability of unequal homologous recombination and illegitimate recombination to generate abundant small deletions and consequently counteract the increase in genome size caused by retroelement insertion (Shirasu *et al.*, 2000; Devos *et al.*, 2002; Feuillet and Keller, 2002; Wicker *et al.*, 2003). Thus, comparative genomics is a powerful approach for the identification of processes involved in genome evolution.

1.4 Intraspecific sequence comparison

Several recent studies have compared sequences of the homoeologous A, B, and D genomes of wheat, a grass genus that includes diploid, tetraploid, and hexaploid species (Figure 3-3). The wheat genomes are estimated to have diverged from a common ancestor between 2.5 and 4.5 million years ago (Huang *et al.*, 2002). Hence, they are ideally suited for comparative studies of recent genome evolution. Different types of the A, B, and D genomes are found in different species. For example, there are two distinct A genomes existing in wheat species: the A^u genome of *Triticum urartu* and the A^m genome in *Triticum monococcum* ssp *monococcum* (einkorn wheat) which is the cultivated diploid A genome species (Figure 3-3). Modern bread wheat (*Triticum aestivum*) is an allohexaploid species that originated from two independent hybridization events. The first one brought together the A^u genome of *Triticum urartu* with the B genome of an unknown species related to *Aegilops speltoides* (SS), resulting in the tetraploid *Triticum turgidum* (AABB) (Figure 3-3; Feldman, 2001). The subsequent hybridization of the early domesticated *T. turgidum* ssp *dicoccum* with the wild species *Aegilops tauschii*, the donor of the D genome, gave then rise to hexaploid wheat *T. aestivum* (AABBDD; Figure 3-3).

Two studies have compared orthologous glutenin gene loci in the A and B genomes of *T. turgidum* ssp *durum* (the cultivated tetraploid wheat), and the D genome of *Aegilops tauschii* (Gu *et al.*, 2004; Kong *et al.*, 2004), as well as in the homoeologous A^u and A^m genomes of *T. durum* and *T. monococcum*, respectively (Wicker *et al.*, 2003). In both cases, conservation between the different genomes was mostly restricted to the gene space whereas intergenic regions were not conserved.

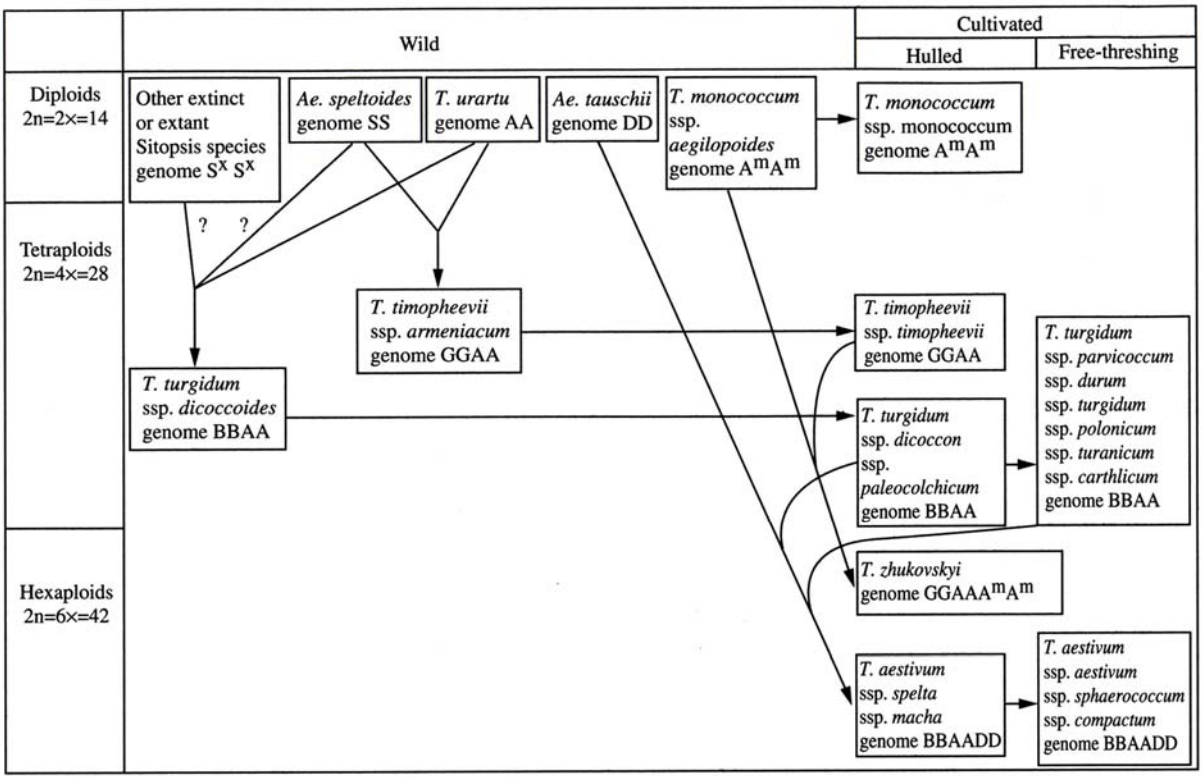


Figure 3-3 Evolutionary relationship of the wheats (Feldman, 2001).

Sequence rearrangements in the intergenic regions were mainly due to the insertions of retrotransposons. Duplications and deletions of large fragments that might be attributed to illegitimate recombination were also observed. Chantret *et al.* (2005) investigated the molecular basis of the evolutionary events that shaped the *Hardness* (*Ha*) locus in the A, B, and D genomes of diploid (*T. monococcum, Ae. tauschii)*, tetraploid (*T. durum*), and hexaploid (*T. aestivum*) wheat species. The *Ha* locus controls grain hardness, an economically important trait of wheat, and comprises the three genes *Pina, Pinb,* and *Gsp-1*. The *Pina* and *Pinb* genes are absent from the A and B genomes of *T. turgidum* subspecies and *T. aestivum*, but present in all their diploid progenitors. Thus, elimination of *Pina* and *Pinb* most likely occurred after tetraploidization. In the D genome of hexaploid wheat, *Pina* and *Pinb* genes were restored upon hybridization of *T. turgidum* ssp *dicoccum* with the D genome donor

Ae. tauschii. The comparative sequence analysis identified a large genomic deletion (that probably occurred independently in the A and B genomes) as the reason for the *Pina* and *Pinb* genes loss from the *Ha* locus of polyploid wheat species. In the B genomes of tetraploid *T. turgidum* and hexaploid *T. aestivum*, the *Ha* locus shows very high sequence conservation (99% identity). In contrast, several genomic rearrangements, such as deletions, inversions, and insertions of transposable elements have differentiated the *Ha* locus in the A and D genomes of diploid, tetraploid, and hexaploid wheat. The authors suggested illegitimate recombination as one of the major mechanisms for these rearrangements (Chantret *et al.*, 2005).

The observed absence of colinearity in these recently diverged species raised the question of sequence rearrangements within different lines or varieties of the same species. A first study, comparing BAC sequences from the *bz1* genomic region of two maize inbred lines (McC and B73), surprisingly revealed as dramatic differences between these two lines as was observed between orthologous loci in two different species (Fu and Dooner, 2002). Not only the composition and location of the intergenic retrotransposon clusters differed, but also the density and content of the genes themselves was different between the two maize inbreds. Four out of the ten identified genes in McC were absent in B73. Comparison of the orthologous loci containing the *zein* storage protein gene cluster *z1C-1* between the two maize inbred lines B73 and BSSS53 revealed similar dramatic variation in gene copy number and insertions of different transposable and retrotransposable elements at different positions (Song and Messing, 2003). In addition, expression analysis of the genes present at the *z1C-1* locus demonstrated that about the same number of genes (six and seven, respectively) are expressed in the two inbred lines. However, only three of the expressed genes were in common between the two lines. Interestingly, the

interval orthologous to the *z1C-1* region of maize was nearly completely conserved in two rice subspecies, a *japonica* and an *indica* variety (Song *et al.*, 2002). Likewise, gene order was found to be completely conserved between two *japonica* cultivars at the *adh1/adh2* locus (Tarchini *et al.*, 2000). Thus, the rice genome seems to have been much more stable since the divergence of the subspecies *japonica* and *indica* (~0.44 mya; Ma and Bennetzen, 2004) than has been the genome of maize.

Recently, Brunner *et al.* (2005) compared DNA sequences from four allelic chromosomal regions in the two maize inbreds Mo17 and B73. Almost 50% of the total sequence analyzed was found to be non-shared between the two inbreds, most of it consisting of LTR-retrotransposons and other mobile elements. However, they also differed considerably in genic sequences. In total, 23 out of 68 putative genes (34%) were present either only in Mo17 or B73. In contrast to the *z1C-1* locus where half of the non-shared sequence is genic due to extensive local duplications affecting the number of *zein* gene copies in each haplotype, no gene duplication was associated with the non-homologous sequences. The non-shared genic segments correspond to fragments of genes (pseudogenes) and are in most cases found in clusters. Within a cluster, the gene fragments tend to be orientated in the same direction. PCR analysis suggested that the genes which are not shared between Mo17 and B73 at the investigated loci are present elsewhere in the maize genome. Moreover, non-shared genes are not at colinear positions in rice whereas shared genes are usually conserved in gene order and location compared to rice. Thus, the non-shared gene fragments seem to originate from insertions rather than from deletions. Only very recently, Morgante *et al.* (2005) identified the pseudogene clusters as part of non-autonomous *Helitrons*, a recently discovered type of eukaryotic transposable elements (Kapitonov and Jurka, 2001). These transposons

appear to have copied and incorporated genic segments from different genomic locations of the host, clustered them together and duplicated these arrangements via a copy-past transposition to non-allelic loci across the maize genome (Morgante *et al.*, 2005; see also chapter VII).

2. Evolution of plant disease resistance (*R*) gene loci

Plants are attacked by a multitude of disease-causing organisms (pathogens) such as bacteria, fungi, viruses, and nematodes. In response to this threat, plants have developed a wide array of defense mechanisms – from structural barriers and pre-formed antimicrobials to inducible defense responses. Moreover, plants have also evolved sophisticated recognition systems to detect proteins produced during infection by specific races of pathogens. This so-called race-specific disease resistance confers in most cases complete resistance against distinct pathogen races.

2.1 Race-specific resistance in plant-pathogen interactions

H. H. Flor, during his pioneering genetic studies of the interaction between flax and flax-rust, formulated the classical gene-for-gene model for disease resistance in plants (1971). According to this model, a plant resistance reaction is triggered when a specific resistance (*R*) gene product interacts with a specific avirulence (*Avr*) gene product from the pathogen. The resistance reaction is often caused by a genetically programmed death of infected cells (the hypersensitive response, HR), as well as tissue reinforcement and local accumulation of antimicrobial compounds at the site of

infection. The gene-for-gene hypothesis was later translated into a receptor-ligand model in which the R protein functions as receptor and the corresponding Avr protein as ligand (Figure 3-4B). Upon binding of the ligand to the receptor, defense responses are triggered. However, the lack of evidence for a direct interaction between most R proteins and their elicitors, together with the identification of plant proteins that directly interact with both Avr and R proteins, have recently led to a new interpretation of the gene-for-gene model: the guard hypothesis (Van der Biezen and Jones, 1998; Van der Hoorn *et al.*, 2002; Figure 3-4C). This model suggests that R proteins guard proteins (termed "guardees") which are targets of Avr proteins. The interaction between the Avr factor and the guardee is sensed by the R proteins that subsequently triggers defense responses. The model implies that R proteins function in recognition complexes containing their guardees.

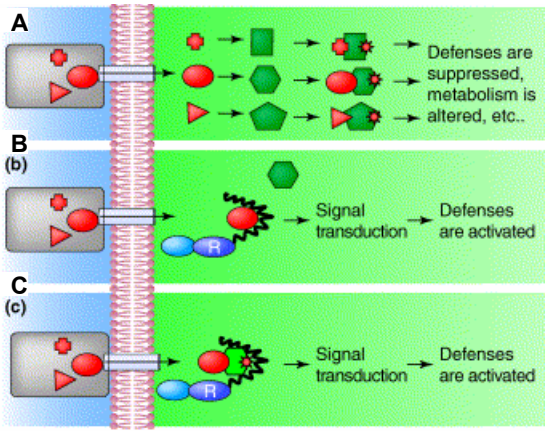


Figure 3-4 Interactions between pathogen Avr proteins and plant R proteins.

A hypothetical pathogen (grey) has attached to a plant cell and is expressing a suite of virulence proteins (red). These proteins are translocated into plant cells via Type III secretion (bacteria) or other unknown mechanisms (fungi and oomycetes). Once inside the cell, they target host proteins (green) that control defense responses, metabolism or other plant process that affect pathogen virulence (note that virulence proteins could also be targeted towards extracellular proteins). (A) In this panel, the plant cell does not express an R protein that is capable of recognizing any virulence protein. Thus, the plant cannot detect the pathogen efficiently and defenses are, at best, only weakly induced. Disease then results from the collective action of the virulence proteins. (B) This panel depicts the classic receptor–elicitor hypothesis, in which an R protein directly binds a virulence protein. This recognition event activates a complex signal transduction network, which in turn triggers defense responses. (C) This panel depicts the guard hypothesis, in which an R protein (guard) detects a modified host protein (guardee, red star), perhaps as a complex with the 'attacking' virulence protein (McDowell and Woffenden, 2003).

2.2 The different classes of *R* genes

Over 40 *R* genes, against different pathogens types, have been cloned and characterized from a variety of plant species (for reviews see Dangl and Jones, 2001; Hulbert *et al.*, 2001; Martin *et al.*, 2003). The majority of the *R* genes can be assigned to one of several different classes based on the combination of a limited number of structural motifs.

By far the largest class of *R* genes known to date encodes proteins with a nucleotide-binding site (NBS) and a series of leucine-rich repeats (LRR). These NBS-LRR proteins have been divided into two major subclasses: those with an amino-terminal TIR (Toll/interleukin receptor) domain (the TIR-NBS-LRR proteins) and those that encode an amino-terminal coiled-coiled motif (the CC-NBS-LRR proteins). In grass species, TIR-NBS-LRR genes have not yet been identified, but the CC-type is very common (Pan *et al.*, 2000).

The other classes of *R* genes are structurally diverse. The rice *Xa21* and *Xa26* genes encode transmembrane receptors with a extracellular LRR domain and an intracellular protein kinase domain (Song *et al.*, 1995; Sun *et al.*, 2004). *Pto* from tomato, which confers resistance against *Pseudomonas syringae,* encodes a serin-threonine kinase (Martin *et al.*, 1993). Another class comprises the *Cf-X* genes from tomato and the Arabidopsis downy mildew resistance gene *RPP27*. They lack an NB site and instead have a transmembrane (TM) domain and an extracellular LRR (*i.e.* receptor-like proteins) (Tor *et al.*, 2004). *RPW8.1* and *RPW8.2*, two related genes that encode proteins with a putative N-terminal TM domain and a CC-motif, confer broad-spectrum resistance to powdery mildew in Arabidopsis (Xiao *et al.*, 2001).

A few R proteins do not fit into one of these classes. The barley stem rust resistance gene *Rpg1* encodes a receptor-like kinase protein with dual kinase

domains (Brueggeman *et al.*, 2002). Another barley *R* gene, *mlo*, is unusual in that the allele conferring broad-spectrum resistance to the barley powdery mildew fungus *Erysiphe graminis* is functionally recessive (Buschges *et al.*, 1997). The deduced protein is predicted to be membrane-anchored and shows similarity to animal G-protein-coupled receptors. Finally, the *Hm1* gene from maize codes for a toxin reductase which confers resistance to a fungal pathogen of maize (Johal and Briggs, 1992) and the Ve proteins from tomato are putative cell-surface glycoproteins with receptor-mediated endocytosis-like signals (Kawchuk *et al.*, 2001).

2.3 Organization of *R* gene loci and their evolution

R gene loci can consist of a single gene with multiple and functionally distinct alleles that recognize different races of a pathogen (*e.g.* the flax *L* locus and the *Rpp13* locus of Arabidopsis), or only one resistant allele (*e.g. Rpm1* and *Rps2* of Arabidopsis). However, classic genetic studies demonstrated that many *R* genes belong to families of genes tightly linked in clusters in plant genomes (reviewed by Hulbert *et al.*, 2001). Consistent with this, genome sequencing confirmed that the majority of NBS-LRR-encoding genes are found in numerous clusters in both *Arabidopsis* and rice (Meyers *et al.*, 2003; Zhou *et al.*, 2004). This clustering is predicted to be a consequence of tandem duplications of paralogous sequences resulting from unequal crossing over (Figure 3-5A). In some cases, heterogeneous gene clusters comprise genes from different sub-families, indicating an origin that is different from tandem duplication. Analysis of the Arabidopsis genome also indicated that numerous small-scale events have duplicated or translocated one or small groups of NBS-LRR genes from clusters to unlinked and probably random locations

in the genome (ectopic duplications; Figure 3-5B) (Richly *et al.*, 2002; Meyers *et al.*, 2003). However, a recent study by Baumgarten *et al.* (2003) concluded that independent duplication of individual NBS-LRR sequences among different chromosomes does not occur frequently. They suggested that most of the genomic relocation of NBS-LRR genes to new chromosomal region are explained by duplication and rearrangement of entire chromosome segments (segmental duplication; Figure 3-5C), rather than from small-scale ectopic duplication events.

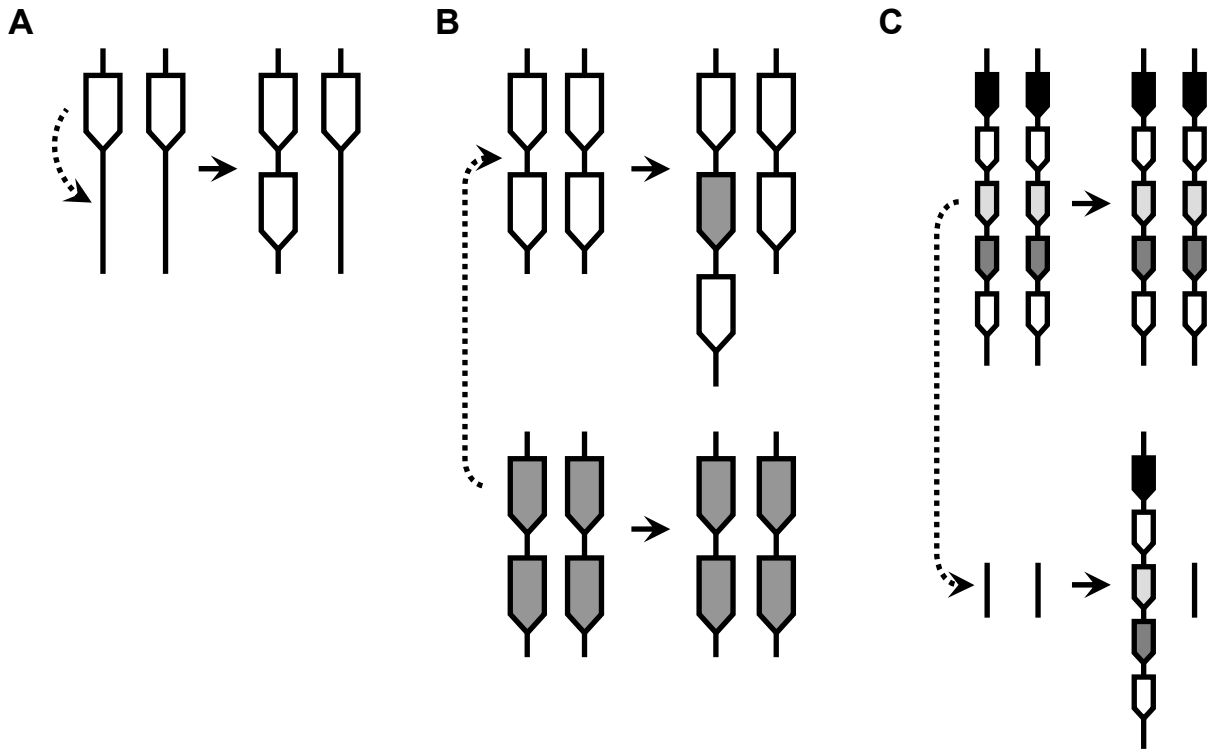


Figure 3-5 Mechanisms contributing to the evolution of *R* gene loci.

(A) Tandem duplication in which the copy is contiguous to the original copy leads to the expansion within a gene cluster. (B) Ectopic duplication transfers individual or small groups of genes to unlinked sites. (C) Segmental duplication involves the duplication and translocation of entire chromosomal regions. (Modified from Leister, 2004).

Colinearity, *i.e.* conserved gene order, is a fundamental feature of grass genomes (see 1.2). However, *R* genes are frequently not present at orthologous loci in different grass species (Leister *et al.*, 1998). For example, although the *Lr10* leaf rust *R* gene from wheat has significant similarity with the powdery mildew *R* genes *Mla1, Mla6, and Mla13* from barley, the positions on the genetic maps of the homeologous chromosomes 1 does not suggest orthology between *Lr10* and the barley genes (Feuillet *et al.*, 2003). A homolog of *Lr10* in barley was identified by Southern hybridization, which maps at the telomeric region of chromosome 1HS where the *Rph4* leaf rust *R* gene is located. However, this gene needs to be mapped in a population segregating for *Rph4* in order to determine whether both, wheat and barley rust genes, are orthologs (Feuillet *et al.*, 2003). Ectopic duplication (Figure 3-5B) has been suggested as a mechanism that relocates individual or small groups of genes to unlinked sites, thereby uncoupling NBS-LRR genes from syntenic relationship (Leister, 2004). Conversely, Baumgarten *et al.* (2003) showed that apparent duplication of NBS-LRR genes to ectopic chromosomal locations is largely the consequence of segmental chromosomal duplication and rearrangements, rather than the independent duplication of individual sequences (Figure 3-5C). When a loss of orthology between closely related species is observed, it might have originated from gene deletion rather than from ectopic events. Intergenic exchange is dramatically lower between NBS-LRR sequences located in different chromosome regions as compared to exchange between sequences within the same chromosome region. Thus, once translocated to new chromosome locations, NBS-LRR gene copies are out of reach of intergenic sequence exchange and are more likely to adopt new functions such as new *R* gene specificities. In this “recombinational isolation”

scenario, physically isolated single-gene loci play a major role in the evolution of *R* gene specificities.

3. Aim of the study

The main objectives of this thesis were to analyze the genomic organization and evolution of the two leaf rust disease resistance loci *Lr10* and *Rph7* in wheat and barley, respectively, and to isolate and functionally analyze the candidate genes for *Rph7* from the resistant barley cultivar Cebada Capa.

In the first part of this work, physical contigs at the *Lr10* locus on chromosome 1AS from diploid (*T. monococcum*), tetraploid (*T. turgidum* ssp *durum*), and hexaploid (*T. aestivum*) wheat were sequenced and analyzed in detail. Sequence comparison of the two haplotypes H1 and H2 defined at the *Lr10* locus revealed that these haplotypes were exceptionally stable throughout evolution and that a large deletion/inversion event is assigned to the origin of the H2 haplotype.

In the second part of the thesis, a physical BAC contig spanning the *Rph7* locus in the resistant barley cultivar Cebada Capa was established and completely sequenced. Comparison of this sequence to the previously sequenced contig from the susceptible cultivar Morex identified seven conserved regions that represent less than 35% of the two sequences. The intervals separating the conserved regions showed completely different sizes as well as numbers and types of repetitive elements.

In a third part of the manuscript I will describe the functional analysis of the four candidate genes for *Rph7* identified in Cebada Capa by *Agrobacterium*-mediated transformation of a susceptible barley cultivar and by virus-induced gene silencing (VIGS) experiments.

IV. ANCIENT HAPLOTYPES RESULTING FROM EXTENSIVE MOLECULAR REARRANGEMENTS IN THE WHEAT A GENOME HAVE BEEN MAINTAINED IN SPECIES OF THREE DIFFERENT PLOIDY LEVELS

Edwige Isidore*, Beatrice Scherrer*, Boulos Chalhoub, Catherine Feuillet, and Beat Keller (2005) Genome Research 15: 526-536

* These two authors contributed equally to this work

1. Abstract

Plant genomes, in particular grass genomes, evolve very rapidly. The closely related A genomes of diploid, tetraploid, and hexaploid wheat are derived from a common ancestor that lived less than 3 million years ago and represent a good model to study molecular mechanisms involved in such rapid evolution. We have sequenced and compared physical contigs at the *Lr10* locus on chromosome 1AS from diploid (211 kb), tetraploid (187 kb), and hexaploid wheat (154 kb). A maximum of 33% of the sequences were conserved between two species. The sequences from diploid and tetraploid wheat shared all the genes, including *Lr10* and *RGA2* and define a first haplotype (H1). The 130 kb intergenic region between *Lr10* and *RGA2* was conserved in size despite its activity as a hot spot for transposon insertion which resulted in more than 70% of sequence divergence. The hexaploid wheat sequence lacks both *Lr10* and *RGA2* genes and defines a second haplotype, H2, which originated from ancient and extensive rearrangements. These rearrangements included insertions of retroelements and transposons, deletions as well as unequal

recombination within elements. Gene disruption in haplotype H2 was caused by a deletion and subsequent large inversion. Gene conservation between H1 haplotypes as well as conservation of rearrangements at the origin of the H2 haplotype at three different ploidy levels indicate that the two haplotypes are ancient and had a stable gene content during evolution, whereas the intergenic regions evolved rapidly. Polyploidization during wheat evolution had no detectable consequences on the structure and evolution of the two haplotypes.

2. Introduction

Comparative analysis of related genomes can reveal the molecular mechanisms of genome evolution. Species from the grass family provide an excellent model for such studies as extensive genetic colinearity among several grass species has been described despite very heterogenous genome sizes and evolutionary divergence times of over 60 million years (for review see Devos and Gale, 2000; Keller and Feuillet, 2000). Recent large scale comparative studies in maize, sorghum, rice, barley and wheat have revealed a mosaic organization of conserved and non-conserved genes at orthologous loci and many small rearrangements (Dubcovsky *et al.*, 2001; Ramakrishna *et al.*, 2002; Song *et al.*, 2002; Brunner *et al.*, 2003; Gu *et al.*, 2003; Ilic *et al.*, 2003; Gu *et al.*, 2004; Guyot *et al.*, 2004). Interestingly, intraspecific comparative studies in maize have also shown disruption of colinearity not only in the intergenic regions but also in the gene space (Fu and Dooner, 2002; Song and Messing, 2003). The partial absence of microcolinearity observed between grass species as well as between inbred maize cultivars is a rich resource for the identification of molecular mechanisms involved in genome evolution.

The wheat A genomes that diverged from a common ancestor living about 0.5-3 million years ago (mya) (Huang *et al.*, 2002; Wicker *et al.*, 2003) are particularly suitable for comparative analysis. The A genomes are found in modern wheat species of different ploidy such as Einkorn wheat (*Triticum monococcum,* diploid), Emmer wheat (*T. turgidum*, tetraploid) and bread wheat species (*T. aestivum*, hexaploid) (Feldman, 2001). A number of recent studies have demonstrated rapid and massive local changes in wheat genomes after polyploidization (Liu *et al.*, 1998;

Ozkan *et al.*, 2001). Therefore, the comparison at the molecular level of wheat A genomes at different ploidy levels could also give insight into the molecular consequences of the “genomic shock” following hybridization of complete genomes. The development of large insert bacterial artificial chromosome (BAC) libraries from diploid wheat *Triticum monococcum* DV92 (AA; Lijavetzky *et al.*, 1999), tetraploid wheat *Triticum turgidum* ssp *durum* cv. Langdon (AABB; Cenci *et al.*, 2003) and the donor of the D genome *Aegilops tauschii* (Moullet *et al.*, 1999) has allowed the isolation and sequencing of large wheat genomic fragments. Recent studies have compared orthologous loci in the three A, B, and D homoeologous wheat genomes (Gu *et al.*, 2004; Kong *et al.*, 2004), which diverged 2.5 to 4 mya (Huang *et al.*, 2002). Rapid genome evolution was observed that was mostly due to insertion of retroelements after the divergence of the three subgenomes. A comparison of the A genome of tetraploid (cv. Langdon with A^u genome from *T. urartu*) with the A^m genome of the diploid *T. monococcum* DV92 that diverged 0.5-3 mya (Huang *et al.*, 2002; Wicker *et al.*, 2003), has recently revealed rapid genome divergence in the intergenic regions even between these closely related genomes (Wicker *et al.*, 2003). Conservation was restricted to small regions and a large proportion of the sequence, which contains mainly repetitive elements, was completely different, providing evidence for a dynamic and rapid evolution. So far, none of the previous comparative studies have included orthologous sequences from hexaploid wheat, and the evolution of A genomes derived from *T. urartu* in different polyploid backgrounds has not been investigated. Two BAC libraries from hexaploid wheat have been recently constructed from the cultivars Chinese Spring (Allouis *et al.*, 2003) and Renan (B. Chalhoub, unpublished data) allowing such analysis.

The leaf rust resistance gene *Lr10* was recently isolated (Feuillet *et al.*, 2003) using a combination of subgenome map-based cloning (Stein *et al.*, 2000; Wicker *et al.*, 2001) and haplotype studies (Scherrer *et al.*, 2002). *Lr10* is closely associated with a second resistance gene analog (*RGA2*) but the two genes are very different from each other at the molecular level (Feuillet *et al.*, 2003). Analysis of the wheat gene pool by Southern hybridization revealed the presence of two groups of lines with distinct haplotypes (H1 and H2) at the *Lr10* locus, defined by the presence (H1) or absence (H2) of the two full-length *Lr10* and *RGA2* genes on chromosome 1A (Scherrer *et al.*, 2002). Thus, the *Lr10* resistance locus differs from most resistance loci in plants, where clusters of closely related genes are present, but is similar to the Arabidopsis *RPM1* locus (Grant *et al.*, 1995). The *RPM1* locus also exists in two stable haplotypes in the Arabidopsis gene pool that are defined by the presence or absence of the gene.

Studies of molecular haplotype diversity in plants are limited and so far restricted to a few species such as Arabidopsis, soybean, barley, and maize (Collins *et al.*, 2001; Nordborg *et al.*, 2002; Charlesworth *et al.*, 2003; Zhu *et al.*, 2003; for review see Rafalski and Morgante, 2004). To assess the molecular basis and mechanisms of genomic rearrangements during evolution of the *Lr10* locus in the two different haplotypes, we have compared BAC sequences at the orthologous *Lr10* loci of the A genomes from diploid, tetraploid and hexaploid wheat species. These comparisons revealed that the two haplotypes were very stable during evolution from diploid to tetraploid and hexaploid wheat, although conservation was restricted to the gene space whereas intergenic sequences diverged rapidly. A large deletion and inversion was found at the origin of the H2 haplotype that, despite gene loss, was highly conserved during wheat evolution.

3. Materials and Methods

3.1 Genomic DNA isolation and PCR

The plant material used in this study is listed in Appendix 9-1A. Genomic DNA isolation was performed as described by Graner *et al.* (1990). The primers used for PCR amplification are listed in Appendix 9-1B and PCR amplifications were performed according to standard procedures (Sambrook *et al.*, 1989), and using the annealing temperatures specific for the different primer pairs.

3.2 BAC clone isolation and sequencing

The *T. turgidum* ssp *durum* Langdon BAC library (Cenci *et al.*, 2003) was screened by hybridization as described by Stein *et al.* (2000) using probes derived from the *T. monococcum* DV92 sequence (Wicker *et al.*, 2001). The *T. aestivum* Renan BAC library was screened by PCR with the primer pair specifically amplifying the *RGA2* copy on chromosome 1D (primers DF: 5'-GTGGTGTTGTGTTGCCAG-3' and DR: 5'-GCAAGCTGGGTCCGGAAC-3') and with a primer pair amplifying *RGA2* from the 1A and 1D chromosomes (primers ADF: 5'-GAAGCCGGATTATAGTGTC-3' and ADR: 5'-CTGCCCAGCTAAGTTCTCT-3') according to standard procedures (Sambrook *et al.*, 1989). BAC DNA preparation for fingerprint analysis and BAC end sequencing were performed as previously described (Stein *et al.*, 2000). Resulting BAC clones from both BAC libraries were assembled into individual contigs based on *Hin*dIII, *Not*I and *Sal*I restriction patterns and hybridization with BAC end probes as well as probes derived from the *T. monococcum* DV92 orthologous sequence (Figure 4-1) (Stein *et al.*, 2000; Scherrer *et al.*, 2002).

Shotgun library construction of the BAC clones 1156G16 and 930H14, and sequencing of shotgun clones for each BAC were performed as follows: The BAC DNA was isolated using the large construct kit (Qiagen). 20 µg of BAC DNA were sonicated and the resulting fragments were purified by agarose gel electrophoresis. The fractions of 1,200-1,800 bp and 2,500-3,000 bp were eluted from the gel and the fragments were subcloned into a *Sma*I-digested pUC19 sequencing vector. The subclones were sequenced using Big-Dye terminator chemistry (Applied Biosystems). Data were collected using ABI 3730 automated sequencers (Applied Biosystems). A total of 700 to 1,200 clones were sequenced for each BAC clone for seven- to ninefold coverage. Gaps between subcontigs were filled by primer walking on the spanning shotgun clones or direct sequencing of BAC DNA. The assembled BAC sequences were finally checked by comparison of *Bam*HI, *Eco*RI, *Hin*dIII, *Not*I and *Pac*I BAC fingerprints with theoretical restriction digests of the assembled sequences.

For the Langdon BAC 1156G16, a total of 2,040 shotgun sequences resulted in 10 contigs with nine gaps. The pre-assembly procedures could not close those gaps due to the presence of G- and C–rich regions or aberrant assemblies caused by multiple long-terminal repeats (LTRs) of similar retroelements nested into each other. Despite the presence of these GC-rich regions, all the gaps could be closed by primer walking. Finally, a total of 187,054 bp of sequence was assembled with a 9.18-fold redundancy for the Langdon 1156G16 BAC clone. For the cv. Renan BAC 930H14, a total of 1,235 shotgun sequences resulted in 12 contigs with 11 gaps caused by similar features to the ones found in the cv. Langdon BAC. A total of 10 of the 11 gaps could be closed by primer walking, leaving one gap of 1 kb as determined from the sizes of the spanning shotgun clones. This 1 kb of missing

sequence was located at base 121554 of the BAC sequence, which is between the *AH5Y* and *NLL2* genes and downstream of the sequence used for the comparative study. We therefore filled it artificially with 1,000 Ns, indicating in the annotation that 1 kb of sequence are missing at the position 121,554 bp of this BAC clone. Finally, a total of 154,778 bp of sequence was assembled. The final error rate for the two BAC sequences was less than 1 base per 10 kb and all finished sequences were of high quality (PHRED scores > 25).

3.3 Sequence analysis

Base calling and quality of the shotgun sequences were processed using Phred (Ewing *et al.*, 1998) and initial assembly of each BAC clone was performed using the Phrap assembly engine (version 0.990319; provided by P. Green and available at http://www.phrap.org). Subcontigs and singlet DNA sequences were analysed using BLAST (Basic Local Alignment Tool) algorithms (Altschul *et al.*, 1997) against public DNA and protein databases. Shotgun clones overlapping between subcontigs were identified using the Staden Package program GAP4 version 2003.0b (Bonfield *et al.*, 1995). Annotation of the genes was performed using a mixture of BLASTX, BLASTN and RiceGAAS annotation system (Sakata *et al.*, 2002). Known repetitive elements were identified by BLAST against TREP (http://www.wheat.pw.usda.gov/ITMI/Repeats; Wicker *et al.*, 2002) database (nucleotide and protein) and dot plot analysis (DOTTER; Sonnhammer and Durbin, 1995). New repetitive elements were identified by dot plot analysis and BLAST against public protein databases. Comparisons of the assembled sequences were performed using dot plot and GCG sequence Analysis Software (Madison, WI). Rates of non-synonymous (*Ka*) vs. synonymous (*Ks*)

nucleotide substitutions (*Ka/Ks*) per 100 sites were calculated for the exon 3 of *RGA2* in *T. monococcum* DV92, cv. Langdon, and cv. Renan with MEGA2 (Kumar *et al.*, 2001) based on the algorithm of Li *et al.* (1993). The same method was applied to calculate the synonymous substitution rates of the other genes. Dating of retrotransposon insertions was performed as described by San Miguel *et al.* (1998) using the MEGA2 software (Kumar *et al.*, 2001). The average substitution rate of repetitive sequences was calculated using the complete conserved repetitive elements between *T. monococcum* and *T. turgidum* and the divergence time of the two loci was estimated as described in San Miguel *et al.* (1998).

4. Results

4.1 Contig establishment at the *Lr10* locus in tetraploid and hexaploid wheat

We have isolated BAC clones from the orthologous *Lr10* loci on chromosome 1AS from *T. turgidum* ssp *durum* cv. Langdon and *T. aestivum* cv. Renan, which belong to the H1 and H2 haplotypes, respectively (Figure 4-1). Seven BAC clones were identified from the tetraploid Langdon BAC library (data not shown) using the probes *rga1NBS* and *rga2NBS2,* which are derived from the *Lr10* and *RGA2* genes respectively (Scherrer *et al.*, 2002) and are specific for the A genome. Previous hybridization studies have shown that H2 lines (including Renan) have no *Lr10* gene, only part of *RGA2* on chromosome 1A and a full copy of *RGA2* on chromosome 1D (Scherrer *et al.*, 2002). No specific primers could be designed for the partial fragment of *RGA2* on chromosome 1A. Therefore, the hexaploid Renan BAC library was screened by PCR of pooled BAC DNA with two primer pairs, one amplifying specifically the *RGA2* copy on the D genome (DF/DR) and the other one amplifying both *RGA2* copies on the A and D genomes (ADF/ADR). In this way, we could identify three BACs from the A genome by screening for amplification with the primers ADF/ADR but not with DF/DR (data not shown). Based on DNA hybridizations with probes derived from the *T. monococcum* DV92 sequence (Wicker *et al.*, 2001), physical contigs were established for the tetraploid and hexaploid wheat species. Hybridization experiments of *Not*I fingerprints of both tetraploid and hexaploid BAC clones revealed that the order of the probes was conserved in both wheat species (Figure 4-1). However, the *Lr10* and *RGA2* probes corresponding to the 5' end of both genes were not detected in the hexaploid Renan (H2 haplotype). The probe *F467* derived from low-pass sequencing of the *T. monococcum* DV92

BAC clone 111I4 (Stein *et al.*, 2000) was found on two *Not*I fragments on the *T. turgidum* BAC suggesting a duplication in cv. Langdon. In addition, the distance between *RGA2* and MWG2245 (Stein *et al.*, 2000) was shorter in the hexaploid H2 haplotype, indicating a large deletion including the *Lr10* gene in Renan. Based on these results, one BAC clone from cv. Langdon (BAC 1156G16) and one BAC clone from cv. Renan (BAC 930H14) was selected for complete sequencing (Figure 4-1). A total of 187,054 bp of sequence was assembled with a 9.18-fold coverage for the Langdon 1156G16 BAC clone. For the cv. Renan BAC 930H14, a total of 154,778 bp of sequence was assembled with a similar coverage.

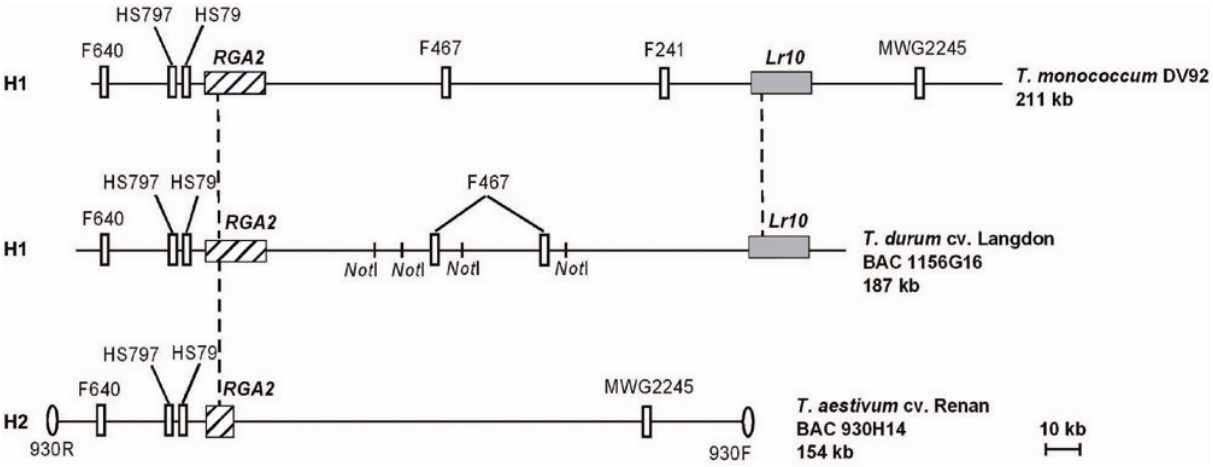


Figure 4-1 Physical maps of BAC clones from the *Lr10* orthologous loci in diploid, tetraploid, and hexaploid wheat.

All probes indicated are derived from the *T. monococcum* DV92 sequence (Wicker *et al.*, 2001) except for 930F and 930R, which are PCR amplified BAC end probes of the hexaploid BAC clone and were not found in any of the other sequences. The *Not*I restriction sites are also indicated.

4.2 Sequence organization at the *Lr10* locus in tetraploid and hexaploid wheat

The tetraploid and hexaploid BAC sequences were annotated based on comparison with the diploid *T. monococcum* DV92 sequence analyzed by Wicker *et al.* (2001). The 187,054 bp sequence of tetraploid wheat is comprised of 8% genic regions and 55.3% identifiable, repetitive elements. Class I LTR retrotransposons are the most abundant elements, and *Angela* LTR retrotransposons represent over 48% of all repetitive sequences. The 154,778 bp sequence of hexaploid wheat is comprised of 11.5% genic regions and 33.3% repetitive elements. *Angela* LTR retrotransposons represent over 50% of all repeats and are mainly located in a single region of approximately 25 kb between the 3' end of *RGA2* (*RGA2-b*) and *NLL1* (Figure 4-2). The only class II elements identified in the three orthologous sequences are CACTA transposons (Wicker *et al.*, 2003). Five genes (*ACT*, *CCF*, *CCF*$^{(p)}$, *RGA2,* and *Lr10*), of which one is a putative gene (*CCF*$^{(p)}$), were found in the Langdon BAC 1156G16 sequence (Figure 4-2). Orthologs of all five genes have been previously identified in *T. monococcum* DV92 (Wicker *et al.*, 2001; Guyot *et al.*, 2004) (Figure 4-2). The *ACT*, *CCF*, and *CCF*$^{(p)}$ genes were also present in the orthologous Renan BAC 930H14 sequence. In this sequence, three additional genes were detected (*NLL1, A5HY,* and *NLL2*) for which orthologs also exist in the orthologous region of *T. monococcum* DV92. The *NLL1* gene was described by Wicker *et al.* (2001) whereas *NLL2* was recently identified by Guyot *et al.* (2004). The *A5HY* gene, which was identified in *T. monococcum* by low-pass sequencing (data not shown), encodes a conserved domain of cytochrome P450s, proteins usually involved in oxidative degradation of various compounds. The best BLASTP hit for this protein is an aldehyde-5-hydroxylase (e value of e-159).

The diploid and tetraploid wheat lines studied here have the same haplotype structure (H1) at the *Lr10* locus as they each comprise full length *RGA2* and *Lr10* genes (Scherrer *et al.*, 2002; Figure 4-2A) as well as the three genes *ACT*, *CCF*, and *CCF*$^{(p)}$ (Figure 4-2). All the genes are conserved in order and orientation. The size of the large intergenic region between *Lr10* and *RGA2* is nearly identical (around 130 kb) in both sequences (Figure 4-2A). In the haplotype H2 of the hexaploid cv. Renan, the genes *ACT*, *CCF, CCF*$^{(p)}$, *A5HY,* and *NLL2* (Figure 4-2B) are conserved in the same order and orientation as in the H1 haplotype. However, major differences are observed in the *RGA2/NLL1* interval. First, in the hexaploid sequence (H2), 2.5 kb of *RGA2,* corresponding to the 5' half of the gene, are missing. The remaining part (bases 2,537 to 4,769) is split into two parts, *RGA2-a* and *RGA2-b,* which are separated from one another by more than 46 kb (Figure 4-2B). In addition, *RGA2-a* is inverted compared to the original gene sequence. The gene fragment *RGA2-a* comprises the last 73 bp of the second exon, the second intron and the first 352 bp of the third and last exon of *RGA2* (bases 2,538 to 3,272 of the complete *RGA2* gene). The gene fragment *RGA2-b* comprises the remaining part of the last exon (bases 3,273 to 4,769 bp). The second major difference to the sequence of the H1 haplotype is the complete absence of *Lr10* in the hexaploid H2 sequence confirming previous hybridization experiments (Scherrer *et al.*, 2002). Finally, the two haplotypes differ in the position of the *NLL1* gene between the *RGA2* fragments, and the inverted orientation of *NLL1* and *RGA2-a* compared to the order observed in the H1 sequences. This suggests that the *T. aestivum* H2 haplotype originated from the H1 haplotype after extensive rearrangements including deletions, inversions, and insertions.

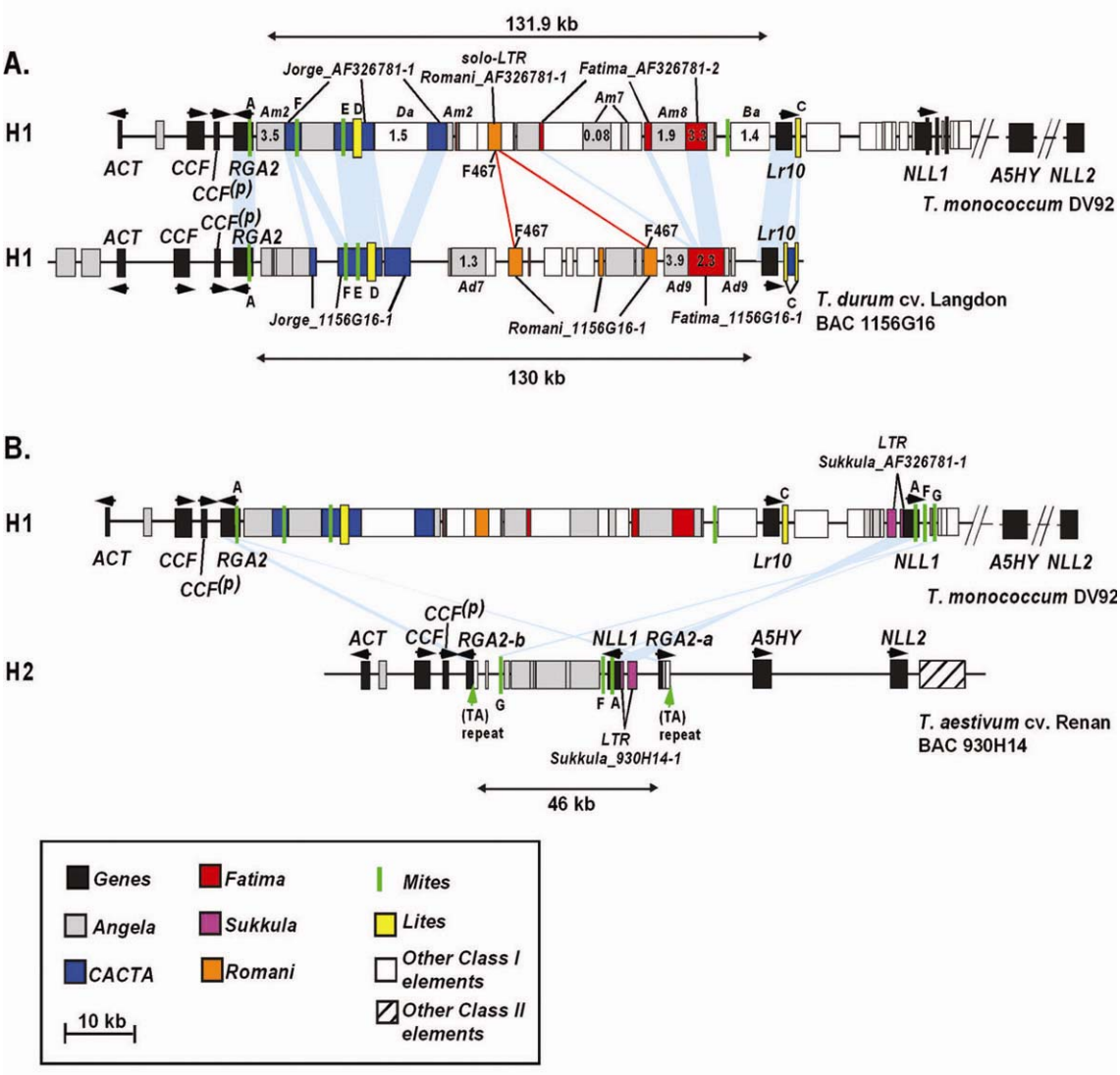


Figure 4-2 Schematic representation of the sequence organization and detailed comparison at the *Lr10* orthologous loci in diploid, tetraploid, and hexaploid wheat.

(A) Organization of the orthologous loci in *T. monococcum* and *T. turgidum* and comparison of the genomic region between *RGA2* and *Lr10* (haplotype H1).
(B) Organization of the orthologous loci in *T. monococcum* and *T. aestivum* and comparison between *T. monococcum* (H1) and *T. aestivum* (H2) in a region delimited by the two *RGA2* fragments of the *T. aestivum* sequence. Light blue areas indicate conservation between the orthologous regions in the *RGA2-Lr10* interval. The genes are indicated by black boxes. The arrowheads located above or below the genes indicate the transcriptional orientation. Different repetitive elements are displayed with different colors. Full names of conserved retroelements are given according to the complete annotation of the BAC sequences AY663391 and AY663392. *Am2*, *Am7*, *Am8*, *Ad7,* and *Ad9* correspond to *Angela* retroelements *Angela*_AF326781-2, *Angela*_AF326781-7, *Angela*_AF326781-8, *Angela*_1156G16-7, *Angela*_1156G16-9, respectively. *Da* and *Ba* correspond to *Daniela*_AF326781-7, *Barbara*_AF326781-7, respectively. The foldback elements (MITES and LITES) are only given

4.3 A gene-rich region is highly conserved in the three homoeologous A genomes

At the three orthologous loci, the first 31 kb of common sequence correspond to a gene-rich region with three genes (*ACT*, *CCF,* and *CCF*$^{(p)}$; Figure 4-3) and a partial gene (*RGA2*) that are conserved in order and orientation. The sizes of the intergenic regions are short, the largest being 14.6 kb between *ACT* and *CCF* in the *T. monococcu*m DV92 sequence (Figure 4-3). Nucleic acid sequences of all the genes are highly conserved, as there is more than 89.8% identity between the orthologous genes independent of haplotype and ploidy level (data not shown). In addition, sequences ranging from 5 bp to over 6.8 kb are also conserved in the intergenic regions with over 80% of identity at the nucleic acid level (Figure 4-3).

The conserved regions downstream of the genes are in general shorter (from 5 bp to 1.3 kb) than the upstream regions (from 28 bp to 6.8 kb). For both the *ACT* and *CCF* genes, the conserved 5' upstream regions are longer in *T. turgidum* and *T. aestivum* (1.3 kb and 6.8 kb, Figure 4-3) than in *T. turgidum*/*T. aestivum* and *T. monococcum* (less than 750 bp in both). In *T. aestivum* (haplotype H2), only the bases 3,273 to 4,769 of *RGA2* (*RGA2-b*), corresponding to the last 1,496 bp of the

Legend to Figure 4-2 continued:

initials in bold type font: A stands for Athos, F for Fortuna, E for Eos, D for Deimos and C for Charon. F467 is a RFLP probe derived from the diploid sequence (Wicker *et al.* 2001) corresponding to an LTR fragment of a *Romani* retroelement and the red lines between the diploid and tetraploid sequences indicate that these LTRs are derived from a common ancestor element. Green arrows indicate the (TA) repeats. Numbers in the boxes indicate the estimated insertion dates of the respective retroelements in million years ago. Numbers below the sequences indicate intergenic distances in kb delimited by double-headed arrows. The sequence in the region delimited by dashed lines at the right end of the *T. monococcum* sequence was only low-pass sequenced, and the intergenic distances in this region were as estimated by Stein *et al.* (2000). H1 and H2 indicate the two haplotypes H1 and H2 according to Scherrer *et al.* (2002).

last exon, are found at the same position as in *T. monococcum* and *T. turgidum* (haplotype H1). Nevertheless, the 1,496 bp of sequence is highly conserved (> 96% of sequence identity). In addition 1.3 kb downstream of the stop codon of *RGA2* are conserved, in all three genomes, with more than 90% identity (Figure 4-3), whereas 4.6 kb are conserved upstream of the start codon of *RGA2* only in the diploid and tetraploid species. A similar conservation of both upstream and downstream regions was found for *Lr10* (data not shown) in the H1 haplotype, and the gene sequence itself is even more highly conserved (over 99% identity). Such conservation in the gene and regulatory sequences of both *Lr10* and *RGA2* indicates a strong selective pressure that maintains them intact and functional in the species of the H1 haplotype. To study whether selective pressure acts on the functional *RGA2* gene and to assess if this is also true for the partial *RGA2* sequence found in the hexaploid wheat cultivar, we have compared the last exon (exon 3) of *RGA2* (nucleotides 2,920 to 4,769) in *T. monococcum*, *T. turgidum* and *T. aestivum*. Estimation of the synonymous (*Ks*) *vs*. non-synonymous (*Ka*) substitution rates yielded a *Ka/Ks* ratio of 0.125, and revealed strong purifying selection on the H1 sequences with a functional allele (*T. monococcum* and *T. turgidum*). In contrast, the *Ka/Ks* ratio of 0.35 calculated for exon 3 in the *T. aestivum* (H2) sequence indicates relaxation of the selection on the partial gene.

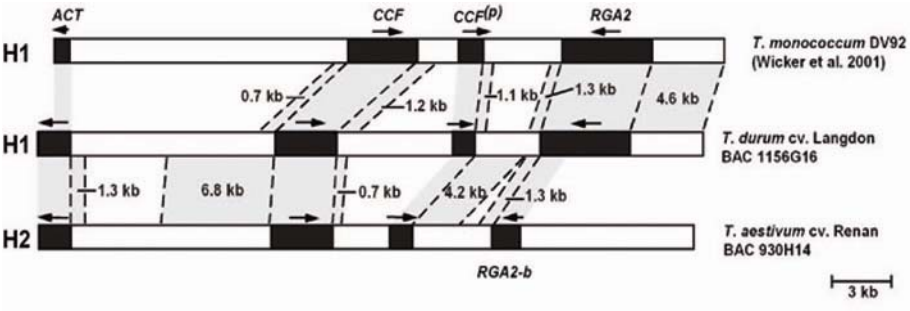


Figure 4-3 Schematic comparison of the 31 kb gene-rich region conserved on the proximal side of the *Lr10* locus in *T. monococcum, T. turgidum,* and *T. aestivum*.

The four genes *ACT*, *CCF*, *CCF(p)* and *RGA2* located in this gene-rich region are represented by black boxes. The arrowheads located above the genes indicate the transcriptional orientation. Areas of similarity between the orthologous sequences are indicated in grey. The sizes of conserved intergenic regions are given in kb and are delimited by dashed lines. H1 and H2 indicate the haplotypes according to Scherrer *et al.* (2002).

4.4 The large intergenic regions between *RGA2* and *Lr10* evolved differentially in *T. monococcum* and *T. turgidum* of the H1 haplotype

The intergenic regions between *RGA2* and *Lr10* in *T. monococcum* and *T. turgidum* are mainly comprised of nested repetitive elements (Figure 4-2A). Repetitive elements represent 64% of the *T. turgidum* interval and 83% of the *T. monococcum* interval, which is higher than the content of repetitive elements in the complete sequences of both species (54% in *T. turgidum* and 70% in *T. monococcum*). This suggests preferential insertion of repetitive elements in this intergenic region. The largest complete elements common to both species are a *Jorge* CACTA transposon of 15.6 kb (*Jorge*_AF326781-1 and *Jorge*_1156G16-1) and a *Fatima* LTR retrotransposon of 9.1 kb (*Fatima*_AF326781-2 and *Fatima*_1156G16-1). These elements are true orthologs as they are found in both species at corresponding positions, indicating that they were present already in the common ancestor. In addition, five foldback elements (three MITES and two LITES,

Figure 4-2A) located in *RGA2*, the *Jorge* element and downstream of *Lr10* respectively, are also conserved in both sequences. These data indicate that these elements inserted between *RGA2* and *Lr10* in the common ancestor of the H1 lines before the divergence of the A genome in the *T. monococcum* and *T. urartu* lineages. Both *Jorge* and *Fatima* elements are interrupted by other elements in diploid wheat but only *Jorge* is interrupted in the tetraploid sequence (Figure 4-2A). In total, 35.7 kb are conserved between the *RGA2* and *Lr10* genes at the two orthologous loci of the H1 haplotype structure. This represents only 27% of the two intergenic regions despite their almost identical length (130 kb).

The presence of the conserved *Jorge* and *Fatima* elements both in *T. monococcum* DV92 and *T. turgidum* cv. Langdon allowed the comparison of base substitution rates in repetitive elements *vs.* genes in this region. The average substitution rate of the complete sequences of the two repetitive elements calculated as described in San Miguel *et al.* (1998) is 0.032. The synonymous substitution rates of the coding sequences of three complete genes (*CCF*, *Lr10,* and *RGA2*) are 0.054, 0.050, and 0.019, respectively. *RGA2* appears to have evolved more slowly than the repetitive elements. However, the data show that the other two genes appear to have evolved faster than the repetitive elements, and highlight the high variability of synonymous substitution rates in different genes. Recently, such a large variability in substitution rates was also found in a set of 24 genes in rice (Ma and Bennetzen, 2004).

Insertion times were estimated for nine complete LTR retroelements in the *RGA2*/*Lr10* intergenic region in the *T. monococcum* and *T. turgidum* sequences using an average base substitution of 6.5×10^{-9} mutations per site per year (Gaut *et al.*, 1996) and following the method used by San Miguel *et al.* (1998). This

revealed a “recent” group (0.08 to 1.5 mya) and a more “ancient” group (2.3 to 3.9 mya) of retrotransposon insertions. The old group is comprised of one element conserved in both sequences (*Fatima*_AF326781-2 and *Fatima*_1156G16-1) and two *Angela* elements (*Am2* and *Ad9*). Although *Am2* is only present in *T. monococcum* and *Ad9* only in *T. turgidum*, both of them are interrupted by conserved elements (*Jorge* and *Fatima*, Figure 4-2A). Therefore, these *Angela* elements were most likely already present in the common ancestor of the two species and subsequently deleted from one of them. Although *Fatima*_AF326781-2 and *Fatima*_1156G16-1 are true orthologs (*i.e.* they inserted in the common ancestor), their calculated insertion times are slightly different (3.3 and 2.3 mya, respectively; Figure 4-2A). This can be explained by the fact that the LTRs of this element are short (490 bp), compared to an average LTR size of 1,438 bp for *Angela* elements, and this resulted in a calculated substitution rate with a higher standard deviation (SD) (in both sequences, SD was at least 0.85 million years). The more recent group of retrotransposon insertions comprises five elements with insertion times ranging from 0.08 to 1.9 mya, four of them in *T. monococcum* (*Da*, *Am7*, *Am8,* and *Ba*) and only one in *T. turgidum* (*Ad7*), with *Am7* being the most recent insertion (0.08 mya). None of these elements are common to both sequences, indicating either that they inserted after the divergence of the wheat A genomes, or that they were present in a common ancestor and were subsequently deleted from one of the two species. The divergence of the wheat A genomes is estimated to have occurred 0.5 to 3 mya ago (Huang *et al.*, 2002; Wicker *et al.*, 2003). Genome divergence is thus in the range of the calculated insertion times of the four remaining “recent” elements (on average 1.5 mya), but there may be differences in evolution rates between LTR sequences and other

sequences (SanMiguel *et al.*, 1998). We cannot, therefore, exclude the possibility that these elements inserted after the species divergence.

Only 27% of sequence in the *RGA2*/*Lr10* intergenic region is conserved between *T. monococcum* and *T. turgidum*. This is the result of independent rearrangements that occurred after the divergence of the wheat A genomes. Comparison of the conserved features identified several molecular mechanisms underlying these rearrangements. In addition to insertions and deletions of LTR retrotransposons, which have also been described in previous studies in wheat (Wicker *et al.*, 2003; Gu *et al.*, 2004), the pattern of conservation of the *Romani* retrotransposons provides interesting insight into the evolution of this interval (Figure 4-2A). Both *T. monococcum* and *T. turgidum* contain a *Romani* element at corresponding positions, indicating that the element was already present in the common ancestor. In *T. monococcum*, only a solo-LTR with identical TSDs (*Romani*_AF326781-1), the result of an intra-element recombination, is present whereas the complete element is present in the tetraploid wheat sequence (*Romani*_1156G16-1, Figure 4-2A). A model for the evolution of this *Romani* element in the two sequences is proposed in Figure 4-4. Interestingly, *Romani* in *T. turgidum* is interrupted by other retroelements in its internal domain. The insertions of these four elements date back 1.8 to 2.8 mya (dated according to San Miguel *et al.,* 1998). It is therefore very likely that at least some of them were already present in the common ancestor of *T. monococcum* and *T. turgidum*. Thus, the intra-element recombination in *T. monococcum* led to the loss of at least 7 kb of genomic DNA (if all elements inserted after the divergence of the two species) and at the most 33 kb (if all elements were present in the common ancestor). The presence of the two LTRs in *T. turgidum* has actually been misinterpreted as a duplication of the

T. monococcum sequence in *T. turgidum* based on hybridization experiments of the tetraploid BAC with the probe *F467* (Figure 4-1), which hybridized to two *Not*I fragments. This probe was derived from the internal sequence of the *Romani* solo-LTR of *T. monococcum* and detected both LTRs in the *T. turgidum* sequence.

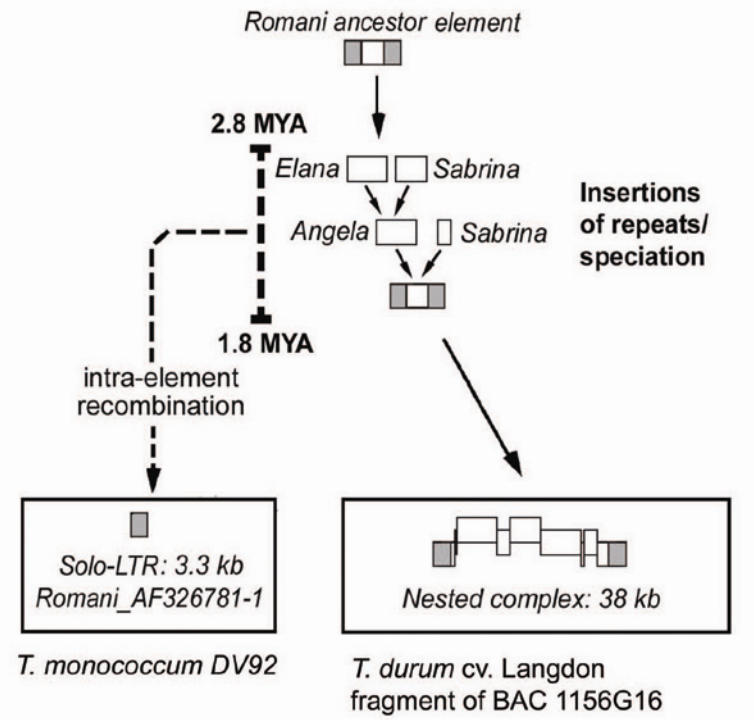


Figure 4-4 Model for the evolution of the *Romani* element in the *T. monococcum* and *T. turgidum* H1 haplotypes.

The complete *Romani* retrotransposon is indicated by rectangles comprising two LTR (dark grey boxes) and internal sequences (white box). The same *Romani* element is at the origin of the solo-LTR in the *T. monococcum* sequence resulting from an unequal intra-element recombination event and at the origin of the nested complex in *T. turgidum*.

4.5 The two haplotypes H1 and H2 originate from ancient and extensive rearrangements

The presence of the conserved *Jorge* and *Fatima* repetitive elements both in *T. monococcum* and *T. turgidum* (a total of 24,797 bp) allowed us to estimate the time of divergence between the two species because these conserved elements must have been identical at the time of divergence. Based on an average base substitution of 6.5×10^{-9} mutations per site per year (Gaut *et al.*, 1996), the two loci diverged approximately 2.4 mya. Similar values for the divergence of the two species were previously obtained by Wicker *et al.* (2003) (2.9 to 3.3 mya). Based on this divergence time and the synonymous substitution rate of *RGA2* between *T. monococcum* and *T. turgidum*, a molecular clock of 3.93×10^{-9} mutations per site per year was established for *RGA2.* Assuming a constant substitution rate in both species, an estimated date of 4 mya was calculated for the partial deletion and disruption of *RGA2* in Renan. Thus, the disruption of *RGA2* in Renan is older than the estimated divergence time of the wheat A genomes (2.4 mya) used in this study. This result indicates an ancient origin of the H2 haplotype and is in agreement with the presence of this haplotype at different ploidy levels.

The longest available sequence of the H1 haplotype (*T. monococcum* DV92, Wicker *et al.,* 2001; Figure 4-2A and 4-2B) was compared to the sequence of the H2 haplotype (*T. aestivum*, Figure 4-2B). The region where the two haplotypes differ dramatically is delimited by the two *RGA2* fragments *RGA2-b* and *RGA2-a* in the *T. aestivum* sequence (50 kb), and by *RGA2* and the foldback element *Gorgon* (G, Figure 4-2B) upstream of the *NLL1* gene in *T. monococcum* (170 kb). On the proximal (left) side of both sequences, *RGA2-b* is conserved in the same orientation (Figure 4-2B). On the distal (right) side, conserved sequences including a fragmented

LTR of a *Sukkula*-type retroelement (*Sukkula*_AF326781-1 and *Sukkula*_930H14-1), the *NLL1* gene, three foldback elements (MITES *Athos* or A, *Fortuna* or F, and *Gorgon* or G) as well as *RGA2-a* are in reverse orientation (Figure 4-2B). Such a complex pattern of conservation and divergence indicates extensive rearrangements at the origin of the two haplotypes. A model for these rearrangements is presented in Figure 4-5, starting with a common ancestor sequence containing *RGA2*, *Lr10*, *NLL1*, *Sukkula,* and the conserved foldback elements. In this ancestral "A" locus, insertions and deletions of repetitive sequences must have occurred between *RGA2* and *Lr10* resulting in an H1 haplotype progenitor which was then further modified to give the modern A^m and A^u haplotype of *T. monococcum* DV92 and *T. turgidum* Langdon (Figure 4-5).

In the evolution of the H2 haplotype, a first deletion event removed a sequence containing the beginning of *RGA2* (corresponding to the first 2,537 bp of *RGA2* of *T. monococcum*) and the complete *Lr10* gene (Figure 4-5). Then, the second half of *RGA2* (bases 2,538 to 4,769) was split into *RGA2-a* and *RGA2-b* at position 3,272 and a large sequence including *RGA2-a*, the foldback elements, the LTR of the *Sukkula*-type element and the *NLL1* gene were inverted (Figure 4-5). The two *RGA2* fragments were disrupted without any sequence loss: at the disruption point, the fragment *RGA2-a* ends at the base 3,272 and the fragment *RGA2-b* starts at the base 3,273. The lack of TSDs or inverted repeats at the breakpoints did not allow the identification of a precise mechanism responsible for this sequence disruption and inversion. Although the same types of fragmented LTR retrotransposons (data not shown) are found in the vicinity of these breakpoints, they do not start exactly at the breakpoints (at least 20 bp after) and are highly degenerated. It is therefore not clear whether they were involved in the rearrangement. Sequence inversion usually

requires inverted repeats that align next to each other and act as recombination sites. Upon cross-over the DNA sequence between these sites is inverted. The only features that could have acted as inverted repeats are (TA) microsatellites of 56 bp and 59 bp found on the hexaploid sequence at the positions 34,333 bp and 85,957 bp. However, their position relative to *RGA2-b* and *RGA2-a* (90 bp and 5 kb downstream respectively, Figure 4-2B) does not allow to identify them as the origin of this rearrangement. Afterwards, insertions and partial deletions involving *Angela* LTR retrotransposons occurred in the inverted sequence, leading to nested *Angela* retrotransposons (Figure 4-2B).

4.6 Identical H2 haplotypes are found at three ploidy levels and H2 sub-haplotypes result from different types of rearrangements

In an earlier study, we did not find the H2 haplotype in tetraploid wheat lines (Scherrer *et al.*, 2002), a finding which contrasted with the large number of H2 lines in hexaploid wheat that evolved from tetraploid wheat (Feldman, 2001). Therefore, we have studied haplotypes at the *Lr10* locus in an additional, large set of 300 tetraploid lines, consisting of 67 accessions of *T. turgidum* ssp *dicoccoides*, 27 *T. turgidum* ssp *carthlicum*, 34 *T. turgidum* ssp *dicoccum*, 20 *T. turgidum* ssp *durum* and 152 accessions of *T. timopheevi* ssp *armeniacum*. A total of 38 of these 300 tetraploid accessions showed the H2 haplotype by Southern hybridization with *Lr10* and *RGA2* probes (data not shown). The 38 lines consisted of eight *T. turgidum* ssp *dicoccoides* accessions, 16 *T. turgidum* ssp *dicoccum*, 2 *T. turgidum* ssp *durum* and 12 *T. timopheevi* ssp *armeniacum* lines, demonstrating the presence of the H2 haplotype in a variety of tetraploid species.

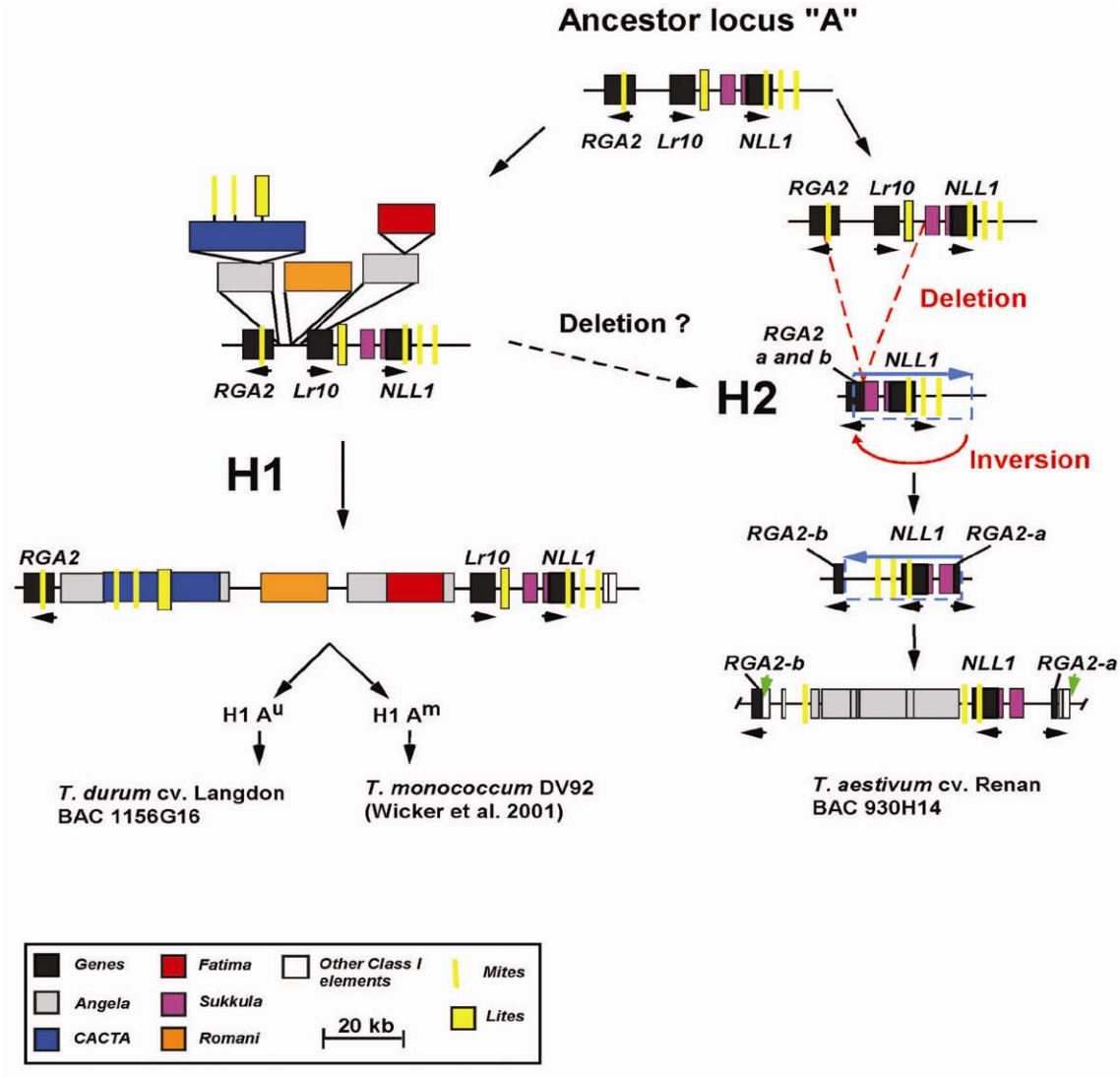


Figure 4-5 Model for the evolution of the H1 and H2 haplotypes at the *Lr10* locus.

Genes are indicated by black boxes. The arrowheads located below the genes indicate the transcriptional orientation. Identified nested repetitive elements are shown by colored boxes. The evolution of the H1 haplotype is represented on the left side and the evolution of the H2 haplotype on the right side of the figure. The ancestor of the H1 haplotype comprises elements common to both *T. monococcum* and *T. turgidum* and subsequent independent events of insertions/deletions occurred resulting in the modern loci. The H2 ancestor has undergone a deletion comprising *Lr10* and the first exon of *RGA2*. The deletion is indicated by red dashed lines. An inversion indicated by a red arrow and delimited by blue arrows and a dashed rectangle also occurred in the H2 ancestor leading to a disruption of *RGA2* that is at the origin of the modern *T. aestivum* sequence. Green arrows indicate the (TA) repeats possibly involved in the sequence inversion.

To assess conservation of the H2 haplotype in the wheat gene pool and to determine whether the same rearrangements are at the origin of diploid, tetraploid, and hexaploid lines with the H2 haplotype, we amplified short genomic fragments (fragments A, B, and C, Figure 4-6A) spanning the deletion and disruption breakpoints of the *RGA2-b* and *RGA2-a* fragments from 56 lines of haplotype H2 (based on hybridization data) using primers derived from the hexaploid sequence (Appendix 9-1). Six diploid *T. monococcum* lines, two diploid *T. urartu* lines (out of 10 tested lines, of which eight had haplotype H1), 10 tetraploid wheat and 38 hexaploid *T. aestivum* lines were analyzed. Based on successful PCR amplification, two groups of lines could be distinguished. The first group amplified all three fragments A, B, and C, like cv. Renan. It comprises all the 38 European hexaploid lines, nine tetraploid lines and four of the six *T. monococcum* lines (*e.g.* lanes 1, 2, 4, 5, 7-11 in Figure 4-6B). The B fragments amplified from *T. monococcum* line TRI 17434 as well as from seven tetraploid wheat lines (see Appendix 9-1) were sequenced. This 331 bp B fragment was identical to the sequence of *T. aestivum* cv. Renan for all lines, demonstrating that the deletion breakpoint is identical. The finding of such a conserved molecular structure strongly suggests that in most lines with the H2 haplotype the deletion derives from the same evolutionary event that occurred in the gene pool of the ancestor species of the wheat A genomes.

A second group of H2 lines comprising two *T. urartu*, two *T. monococcum* and one tetraploid line did not amplify any of the fragments (Appendix 9-1, lanes 3 and 6, Figure 4-6B). Thus, whereas the sequence organization observed in Renan is conserved in all but one of the tetraploid and hexaploid lines of the H2 haplotype, this is not the case for all diploid lines of this haplotype. The presence of this second group of lines suggests either sequence divergence after the formation of the

haplotype from a common ancestor or completely independent deletion events leading to similar haplotype structures. To further investigate conservation among the 51 H2 lines which had the same PCR amplification pattern as Renan, we amplified a 3 kb fragment spanning the deletion breakpoint of the *RGA2-a* fragment with a primer common to the B fragment previously described and a primer located 3 kb upstream in the sequence (Figure 4-6A). Successful PCR amplification was observed for 40 lines (three *T. monococcum,* seven tetraploid lines*,* and 30 *T. aestivum* lines) whereas no amplification was observed in 11 lines (one *T. monococcum,* two tetraploid, and eight *T. aestivum* lines, Appendix 9-1) suggesting sequence divergence in these lines. In conclusion, we could classify 56 H2 lines into three sub-H2 haplotypes. The presence of a main haplotype, represented by Renan, including diploid *T. monococcum*, tetraploid and hexaploid lines, indicates high conservation and stability of the H2 haplotype during evolution and throughout polyploidization events in the last 1-3 million years of A genome divergence.

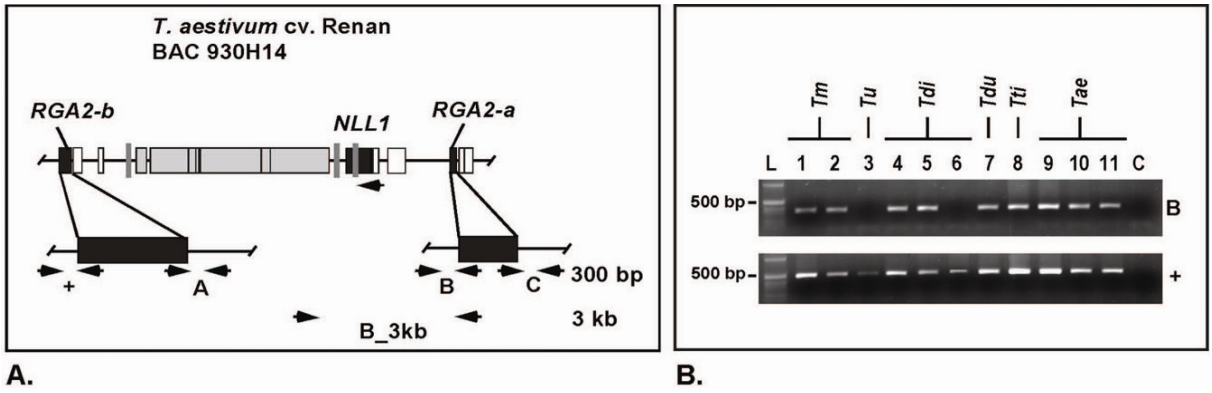


Figure 4-6 Analysis of H2 haplotype conservation in the wheat gene pool.

(A) Schematic representation of the amplified fragments spanning the *RGA2* fragments of *T. aestivum*. (B) PCR amplification of 331 bp fragment [B] and the positive control for amplification [+]. PCR fragments span the 3' end of *RGA2* for the amplification control and the deletion breakpoint for the B fragment. Two *T. monococcum* lines (*Tm*, lanes 1 and 2, TRI 2006 and TRI 17434), one *T. urartu* line (*Tu*, lane 3, TRI 17413), three *T. turgidum* ssp *dicoccoides* lines (*Tdi*, lanes 4, 5 and 6, TRI 3425, TTD22, IG 46516), one *T. turgidum* ssp *durum* line (*Tdu*, lane 7, TRI 673), one *T. timopheevii* ssp *armeniacum* line (*Tti*, lane 8, TA 952), and three *T. aestivum* lines (*Tae*, lanes 9 to 11, cvs. Chinese spring, Arina and Renan) of the haplotype H2 are shown. L indicates 1 kb ladder and C water control.

5. Discussion

We have compared at the molecular level the organization and evolution of two haplotypes at the *Lr10* locus on the wheat A genomes of three different ploidy levels. The recent construction of two BAC libraries from hexaploid wheat lines belonging to the H2 haplotype (cv. Chinese Spring and Renan) (Allouis *et al.*, 2003, unpublished data) allowed the molecular comparison of the two haplotypes, including for the first time a sequence from hexaploid wheat. Conservation between the two haplotypes was limited to the genes located on both sides of a large 150 kb region including the *Lr10* and *RGA2* genes, whereas extensive rearrangements occurred within this region.

Sequence comparison between H1 haplotypes derived from two different species revealed complete conservation of the structure and orientation of all the genes whereas the composition of the intergenic regions, which mainly consist of blocks of repetitive elements, is very different. Strikingly, despite these differences, the length of the large intergenic region of about 130 kb between the *RGA2* and *Lr10* genes is very similar. Conservation of the distances between blocks of genes has also been observed at the *Glu-3A* locus (Wicker *et al.*, 2003) as well as in two maize cultivars at the orthologous *bz*1 loci (Fu and Dooner, 2002), although the intergenic regions were extensively rearranged. This suggests that the specific length (and not necessarily content) of an LTR retrotransposon cluster is conserved during evolution of orthologous loci, possibly acting as a determinant of chromatin and chromosome structure (Bennetzen and Ramakrishna, 2002).

5.1 A large deletion/inversion event is at the origin of the H2 haplotype

In the grass genomes, gene loss plays an important role in genome evolution and is the basis of the mosaic conservation of orthologous sequences (*e.g.* Song *et al.*, 2002; Ilic *et al.*, 2003). The *Lr10* haplotype evolution provides an interesting example where the molecular events leading to such a gene loss could be studied in detail. The extensive rearrangements found at the origin of haplotype H2 are due to a large deletion that includes the disease resistance gene *Lr10* and part of *RGA2.* The molecular basis of this deletion is possibly an illegitimate recombination event, a mechanism that is at the origin of other rearrangements in wheat (Wicker *et al.*, 2003; Ma *et al.*, 2004). So far, the only mechanism responsible for gene disruption that has been described in wheat was the insertion of retroelements into genes (Gu *et al.*, 2004). In the H2 haplotype, the deletion was followed by a large inversion that led to the disruption of the remaining *RGA2* gene fragment. The molecular mechanism of this sequence inversion could not be determined, but the TA repeats found on both sides of the inverted fragment might be relevant. It is likely that during the estimated 4 million years since the deletion/inversion events all tracks of the original sequences have been covered by other elements. Sequence inversions have already been described in other comparative studies in grasses but they specifically concerned complete genes in the distantly related barley and rice species (Dubcovsky *et al.*, 2001) or groups of genes in rice compared to sorghum and maize (Bennetzen and Ramakrishna, 2002). Interestingly, Dubcovsky *et al.* (2001) have also identified two sequences of inverted homology in the vicinity of gene 2, which was found inverted in barley compared to rice. In this case, the inversion did not cause disruption of the gene.

The deletion/inversion events in the H2 haplotype had several consequences for the further evolution of the *Lr10-RGA2* region in the wheat gene pool. In fact, it may have effectively inhibited recombination between the different haplotypes. If recombination occurs within the inverted sequence (*e.g.* in or near the *NLL1* gene), the recombinant gametes will carry chromosomes that are partially deleted or duplicated. Such gametes probably have a greatly reduced fitness and recombination between the haplotypes in the *Lr10-RGA2* interval is effectively suppressed. This provides a molecular explanation for two earlier observations: first, there are no recombinant haplotypes between H1 and H2 in the gene pool and the *Lr10* gene was always found together with an intact *RGA2* (Scherrer *et al.*, 2002). Second, there was no recombination between the two genes in a segregating population of more than 6,000 gametes originating from a cross between the cultivars Frisal and Thatcher*Lr10* (Stein *et al.*, 2000), which belong to the H2 and H1 haplotypes, respectively. Thus, our data indicate that recombination in hexaploid wheat can not only be blocked by chromosomal segments derived from recent introgressions from wild relatives, but also by inversions which originate from ancient haplotypes.

5.2 Common themes of evolution in both haplotypes

Insertions of retroelements followed by deletions, most likely through illegitimate or unequal recombination, played a prominent role in the evolution of both haplotypes, similar to observations in other wheat comparative studies (Wicker *et al.*, 2003; Gu *et al.*, 2004). These two mechanisms were also found to be responsible for LTR retrotransposon removal in the Arabidopsis and rice genomes (Devos *et al.*, 2002; Ma *et al.*, 2004) and seem to play a key role in plant genome evolution in general. Dating of intact elements in *T. monococcum* and *T. turgidum* revealed two

distinct groups of insertion times for LTR retrotransposons. It is possible that different types of retroelements evolve at different rates and that dates of retroelement insertions may in general be overestimated (SanMiguel *et al.*, 1998; Ma *et al.*, 2004). However, “old” and “recent” *Angela* retroelements, which are likely to evolve at the same rate, are found at the same locus supporting the idea that two series of insertions of repetitive elements have occurred at the *Lr10* locus in both diploid and tetraploid species. This suggests that at least two waves of retrotransposon insertions have occurred during the evolution of the A ancestral wheat genome.

It is very rare to find conserved sequences of transposable elements in colinear regions of cereal species and, so far, only partial sequences of conserved elements have been reported (Wicker *et al.*, 2003; Gu *et al.*, 2004; Kong *et al.*, 2004). The half-life of LTR retrotransposons was recently estimated to be approximately 6 million years in rice (Ma *et al.*, 2004). Therefore, the conservation of nine complete elements (including one CACTA transposon, one LTR retrotransposon, one LITE, and 6 MITES) at the *Lr10* locus is exceptional and might be due to a lower tolerance of random rearrangements in this region of the genome. These 25 kb of conserved sequences from repetitive elements allowed us to estimate a divergence date between the wheat A genomes which had previously been estimated to 0.5 to 3 mya (Huang *et al.*, 2002; Wicker *et al.*, 2003). Our estimate (2.4 mya) confirms at a larger scale what had previously been published by Wicker *et al.* (2003) based on 8 kb of conserved sequence. However, it differs from the results of Huang *et al.* (2002). This is probably due to variations in rates of sequence divergence (*e.g.* study of genes *vs*. repetitive elements). Such variations in substitution rates have been found at different levels. Repetitive elements may diverge more rapidly than genes, as shown in rice (Ma and Bennetzen, 2004). Ma and Bennetzen (2004) have also shown that the two

rice subspecies *indica* and *japonica* have diverged at different rates and they suggest that, even within a genome, separate regions may evolve at different rates. Finally, genes have very different divergence rates (Ma and Bennetzen, 2004). All these variations in rates of sequence divergence between subspecies, within a genome at different loci or even among genes of the same locus are very intriguing and seem to be common in grasses. Further sequence comparisons between and within grass species, or even wheat varieties are needed at a larger scale to deepen our understanding of these variations.

The conservation of two ancient haplotypes in wheat species of three ploidy levels suggests that no significant rearrangements have affected the *Lr10* locus during the polyploidization events that are at the origin of tetraploid and hexaploid wheat. Thus, there is no evidence in this region for rapid genomic changes following polyploidization as observed for a substantial proportion of non-coding regions in newly synthesized allopolyploids (Liu *et al.*, 1998; Ozkan *et al.*, 2001). The stability of the two haplotypes at the *Lr10* locus throughout evolution and polyploidization is striking and contrasts with the assumption that disease resistance loci are generally more variable than other loci. This might only be true for some *R* gene clusters, but not for loci under balancing selection (see below).

5.3 Old and stable haplotypes at the *Lr10* resistance locus

The arms race model for host-pathogen interactions predicts resistance loci to be young and monomorphic. Neither is true for the *Lr10* locus, making this wheat resistance locus a new member of a group of resistance loci described in Arabidopsis and tomato that existed before speciation and are also old and polymorphic (Grant *et al.*, 1998; Stahl *et al.*, 1999; Bergelson *et al.*, 2001; Riely and Martin, 2001).

Similar to Arabidopsis *RPM1*, which confers resistance to *Pseudomonas syringae*, there is a presence/absence polymorphism for *Lr10* in the wheat gene pool. This H2 haplotype caused by the deletion event is identical in lines of all three ploidy levels as determined by sequencing the deletion breakpoint in the *RGA2* gene. Thus, the polymorphism represented by the two haplotypes is evolutionarily stable and the two haplotypes have coexisted for 4 million years, suggesting a balanced polymorphism. The *Lr10-RGA2* complex might confer a fitness advantage under certain environmental conditions whereas fitness costs are associated with it under other conditions, resulting in natural selection for both haplotypes and their maintenance in the wheat gene pool. In the absence of the pathogen, fitness costs associated with a resistance gene have been shown for the *RPM1* gene (Tian *et al.*, 2003). Evidence for balancing selection at a resistance locus was described for the Arabidopsis *RPS5* gene (Tian *et al.*, 2002). There, an island of enhanced sequence variability could be detected around the gene, indicative of an old polymorphism under selection.

Our study of the *Lr10* locus demonstrates the power of comparative studies between the A genomes of different ploidy levels in wheat to unravel molecular mechanisms involved in genome evolution. It has also provided a molecular explanation for previous observations at the genetic and phenotypic levels. The detailed molecular understanding of polymorphism at this resistance locus, including its consequence on recombination, provides a solid basis for further evolutionary studies. The presence of both haplotypes in the large collection of accessions of wild diploid and tetraploid wheat in the gene banks will allow future studies correlatingeco-geographical and molecular data for a better understanding of the environmental factors underlying balancing selection and haplotype stability.

V. LARGE INTRASPECIFIC HAPLOTYPE VARIABILITY AT THE *RPH7* LOCUS RESULTS FROM RAPID AND RECENT DIVERGENCE IN THE BARLEY GENOME

Beatrice Scherrer, Edwige Isidore, Patricia Klein, Jeong-soon Kim, Arnaud Bellec, Boulos Chalhoub, Beat Keller, and Catherine Feuillet (2005) The Plant Cell 17: 361-374

1. Abstract

To study genome evolution and diversity in barley (*Hordeum vulgare*), we have sequenced and compared more than 300 kb of sequence spanning the *Rph7* leaf rust disease resistance gene in two barley cultivars. Colinearity was restricted to five genic and two intergenic regions representing less than 35% of the two sequences. In each interval separating the seven conserved regions, the number and type of repetitive elements were completely different between the two homologous sequences, and a single gene was absent in one cultivar. In both cultivars, the non-conserved regions consisted of ~53% repetitive sequences mainly represented by long-terminal repeat retrotransposons that have inserted less than 1 million years ago. PCR-based analysis of intergenic regions at the *Rph7* locus and at three other independent loci in 41 *H. vulgare* lines indicated large haplotype variability in the cultivated barley gene pool. Together, our data indicate rapid and recent divergence at homologous loci in the genome of *H. vulgare,* possibly providing the molecular mechanism for the generation of high diversity in the barley gene pool. Finally, comparative analysis of the gene composition in barley, wheat (*Triticum aestivum*), rice (*Oryza sativa*), and sorghum (*Sorghum bicolor*) suggested massive gene movements at the *Rph7* locus in the Triticeae lineage.

2. Introduction

Comparative genetic mapping in grasses has shown that the gene order is well conserved along the grass chromosomes, despite large differences in genome size, chromosome number, ploidy level, and content of repetitive elements (Gale and Devos, 1998; Bennetzen, 2000; Keller and Feuillet, 2000; Gaut, 2002). Recently, comparative studies of large stretches of BAC DNA sequences at orthologous loci in barley (*Hordeum vulgare*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*), maize (*Zea mays*), and rice (*Oryza sativa*) have shown numerous exceptions to colinearity at the molecular level, revealing a mosaic gene conservation in the grass genomes that was overlooked by genetic mapping. These studies have shown that rearrangements caused by insertion/deletion of transposable elements, duplication, insertion, deletion, and inversion of genes as well as gene movements have shaped the different grass genomes since their divergence from a common ancestor 50 to 70 million years ago (mya) (for recent reviews, see Feuillet and Keller, 2002; Bennetzen and Ma, 2003). Comparative genomic analysis also provided insight into the dynamics of genome evolution by revealing some of the mechanisms underlying the rearrangements such as the rapid turnover of long-terminal repeat (LTR) retrotransposons by unequal and illegitimate recombination or gene amplification followed by gene movements at hotspot regions for sequence divergence (Bennetzen, 2002; Li and Gill, 2002; Ramakrishna *et al.*, 2002; Song *et al.*, 2002). Finally, these studies have revealed differences in the dynamics of genome evolution in different lineages of the grass family and have demonstrated the importance of whole or partial genome duplications during the evolution of these genomes (Gaut *et al.*, 2000; Ilic *et al.*, 2003; Paterson *et al.*, 2003; Guyot and Keller, 2004).

To date, most of the studies have compared orthologous loci between members of the grass subfamilies Andropogoneae (maize and sorghum), Pooideae (barley and wheat), and Erhartoideae (rice). Comparisons between and/or within these families have provided information on rearrangements that have occurred in a time frame of 50 to 60 to 10 to 14 million years (Bennetzen and Ramakrishna, 2002; Feuillet and Keller, 2002; Ramakrishna *et al.*, 2002; Song *et al.*, 2002; Bennetzen and Ma, 2003; Gu *et al.*, 2003; Ilic *et al.*, 2003). More recently, comparisons have been performed between the homoeologous genomes of wheat that have radiated 2.5 to 4.5 mya (Huang *et al.*, 2002). Large sequences were compared at orthologous storage protein loci in the homoeologous A and A^m genomes of *T. monococcum* and *T. durum* (Wicker *et al.*, 2003) and in the A, B, and D genomes of *T. durum* and *Aegilops tauschii* (Gu *et al.*, 2004; Kong *et al.*, 2004). In both cases, conservation was mainly restricted to the gene space, and rearrangements, including gene duplications and rapid divergence of intergenic regions through the movement of retroelements, were observed. Thus, even in recently diverged species, rapid and dynamic genome evolution limits the possibilities to assess the type, rate, and precise mechanisms of microrearrangements in grass genomes (Wicker *et al.*, 2003). Intraspecific comparisons were expected to help address these questions. Very recently, Fu and Dooner (2002) and Song and Messing (2003) have compared BAC sequences from different inbred lines of maize. Surprisingly, there was as much rearrangement between two homologous regions in maize as between orthologous loci in two different species. Dramatic differences were not only found in the composition and length of retroelement blocks in the intergenic regions but also in the gene space where several genes were missing. Fu and Dooner (2002) suggested that gene deletions might have resulted from the retrotransposon invasion that

occurred in the last 2 to 3 million years in the maize lineage. Moreover, they postulated that independent events of retrotransposon invasion in different individuals of the population of modern maize progenitors are at the origin of the high intergenic sequence variability found in maize. These findings raised many questions. Is this exceptional haplotype variation restricted to the maize genome history, and does it only concern particular regions of the genome, such as those carrying genes that are not essential for plant development (*e.g.* pigmentation genes)?

We have recently sequenced 212 kb at the leaf rust resistance locus *Rph7* in the susceptible barley cultivar Morex (Brunner *et al.*, 2003). Here, we have isolated a physical contig of 350 kb from the homologous region in the resistant cultivar Cebada Capa and have compared 226 kb of this contig with 126 kb of sequence from Morex. No conservation in the size and composition of the intergenic regions and the absence of one gene in Cebada Capa indicate rapid and recent genome divergence in the barley genome. In addition, PCR-based haplotype analyses at the *Rph7* locus and at three other independent loci revealed large variability in the intergenic regions in the cultivated *H. vulgare* gene pool. Finally, comparative analysis of the gene composition at the *Rph7* locus in four grass genomes indicates gene movements specifically in the Triticeae lineage.

3. Materials and Methods

3.1 Plant material

The characteristics and the origin of the 41 barley (*Hordeum vulgare*) breeding lines used in the PCR-based haplotype analysis are described by Brunner *et al.* (2000). Among these lines, 14 are known to possess the *Rph7* resistance allele (Table 5-2). The other lines either have other *Rph* genes (Weibull *et al.*, 2003) or are not known to contain any leaf rust resistance gene (Table 5-2).

3.2 Shotgun sequencing and sequence analysis

Screening of the Cebada Capa BAC library and identification of single BAC clones was performed as described by Isidore *et al.* (2005). PCR screening for the *Hvgad1* gene was performed with the following primers: gad-1 (5'-CACACCACGCC-TACTCCTAC-3') and gad-2 (5'-ACGAAGGACGCCAGGTTCAG-3'). Preparation of BAC DNA for fingerprint analysis and BAC-end sequencing as well as shotgun libraries for sequencing the BAC clones 14E11 and 124A1 was made as previously described (Stein *et al.*, 2000). A total of 3,779 clones were sequenced on an ABI PRISM 377 automatic sequencer (Applied Biosystems, Foster City, CA) from both ends (average length, 788 bp). Base calling and quality of the shotgun sequences (predicted error rate 1/3 kb) were processed using PHRED (Ewing *et al.*, 1998) and assembled using the PHRAP assembly engine (version 0.990319; provided by P. Green, http://www.phrap.org). The STADEN software package was used to finish the assembly (Bonfield *et al.*, 1995). Gaps between the subcontigs were filled by PCR using 18- to 24-mer oligonucleotides designed at the contig ends. DNA sequences were analyzed using BLASTN, BLASTX, and TBLASTX algorithms (Altschul *et al.*,

1997) against public DNA and protein sequence databases. Detailed sequence analysis was performed with the GCG sequence analysis software package version 10.1 (Madison, WI) and by dot plot analysis (DOTTER; Sonnhammer and Durbin, 1995). Analysis of repetitive sequences and transposable elements was performed by BLASTN and BLASTX searches against public databases, the database for Triticeae repetitive DNA (Wicker *et al.*, 2002), and a local database for repetitive DNA. For gene prediction, the RiceGAAS annotation system (Sakata *et al.*, 2002) as well as comparison against the rice (*Oryza sativa*) full-length cDNA collection (Kikuchi *et al.*, 2003) was used. BLASTN and TBLASTX searches against the 12 rice pseudomolecules recently released by TIGR [http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml] and against the mapped wheat (*Triticum aestivum*) EST database [http://wheat.pw.usda.gov/wEST/blast/] were performed to identify the position of genes orthologous to the barley genes in rice and wheat, respectively.

3.3 Identification and mapping of sorghum BACs containing genes homologous to the *Rph7* locus

Sequences of the barley genes *Hvgad1, HvHGA1, HvHGA2, Hvpg4, Hvpg1*, and *Hvpg3* were used to identify homologous sequences in the sorghum (*Sorghum bicolor*) EST collection [http://fungen.org/blast/blast.html]. STS primers were designed against each sorghum EST using primer3 software [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi] and used in PCR reactions to screen six-dimensional BAC DNA pools from the sorghum genotype IS3620C (Klein *et al.*, 2003). Positive BACs were identified after electrophoresis on 4% agarose gels and visualization with SYBR Gold (Molecular Probes, Eugene, OR). BAC clones were subjected to low-pass sequence scanning to assess gene content and to aid in

alignment to the rice genome sequence as described by Klein *et al.* (2003). BACs were fingerprinted using high information content fingerprinting (Klein *et al.*, 2003) and integrated into the sorghum physical map with FPC V7.1 (Soderlund *et al.*, 2000). A subset of the positive BACs (76H6, 67F2, 73H13, and 62H18) integrated into existing DNA contigs that had been previously linked to the sorghum genetic map (Klein *et al.*, 2000). FISH was used to map the chromosomal location of BACs 52N1, 61E22, and 63A11, containing a homolog of *Hvpg1*. Cells for chromosome spreads were prepared from anthers of sorghum genotype BTx623, and BAC DNA used for FISH was isolated as described by Islam-Faridi *et al.* (2002). For BAC mapping, a probe cocktail containing the labelled BAC DNAs (52N1, 63A11, and 61E22) as well as two BACs developed previously as karyotype probes for sorghum chromosome 5 (Kim *et al.*, 2002) were hybridized to sorghum pachytene spreads according to Hanson *et al.* (1996). To assess the location and relative intensity of FISH signals, blue (4',6-diaminosino-2-2-phenylindole signal from chromosomal DNA), green (fluorescein isothiocyanate from BAC probes), and red (Cy3 from BAC probes) signals were measured from digital images using Optimas v 6.0 (Media Cybernetics, Carlsbad, CA). Image capture and data analysis were done as described (Islam-Faridi *et al.*, 2002).

3.4 Dating of retrotransposon insertions

Dating of LTR retrotransposon insertions and divergence time estimations were performed based on the method described by SanMiguel *et al.* (1998). MEGA2 (Kumar *et al.*, 2001) was used to calculate the number of transition and transversion mutations. Insertion dates were estimated using the Kimura two-parameter method

(Kimura, 1980). A mutation rate of 6.5 x 10^{-9} substitutions per synonymous site per year, based on the *adh1* and *adh2* loci of grasses (Gaut *et al.*, 1996), was applied.

3.5 Haplotype analysis

PCR was performed on 50 ng of genomic DNA in 25 µL containing 0.5 units of *Taq* DNA-polymerase (Sigma-Aldrich, Buchs, Switzerland), 13 PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.001% gelatin), 100 mM deoxynucleotide triphosphate, and 400 nM primers described in Appendix 9-5. Amplifications were performed in a PTC-200 thermocycler (MJ Research, Bioconcept, Switzerland) as follows: 3 min at 94°C, 30 cycles of 45 s at 94°C, and 45 s at 53 to 63°C depending on the primer combination (Appendix 9-5) followed by 1 min (CR4-3/CR5-1 and CR3-1/CR4-1) or 2 min (CR4-3/CR5-2) at 72°C. The extension of the amplified products was achieved at 72°C for 5 min. PCR with the primers pairs CR4-2/CR5-1, 635P2_G4F/635P2_G5R, and 615K1_G2F/615K1_G3R was performed on 50 ng of genomic DNA with the *TaKaRa ExTaq* polymerase (TaKaRa, Dalian, Japan) according to the manufacturer's instructions. A control PCR was performed with the ITS14/ITS15 primers that amplify a 240 bp fragment corresponding to the rDNA internal transcribed spacer region, ITS1, as described by De Bustos *et al.* (2002). PCR products were separated by electrophoresis on 0.75 to 1.5% agarose gels and visualized under UV light after ethidium bromide staining.

4. Results

4.1 Establishment of a 350-kb Contig at the *Rph7* Locus in Cepada Capa

We have recently sequenced 212 kb at the *Rph7* locus in the leaf rust susceptible barley cultivar Morex (Brunner *et al.*, 2003). Ten putative genes were identified, whereof five are candidates for *Rph7*. None of them showed characteristics of known disease resistance (*R*) genes, suggesting that *Rph7* is either a new type of *R* gene or that it is absent in cv. Morex. To test these hypotheses, we have constructed and screened a pooled BAC library of the *Rph7* donor line, cv. Cebada Capa (Isidore *et al.*, 2005). Screening was performed by PCR with primers corresponding to the *Hvpg1* and *Hvpg4* genes, which cosegregate with *Rph7*, as well as to the *Hvgad1* gene, which is located proximal to *Rph7* (Figure 5-1). In total, 10 independent BAC clones were isolated. Hybridization of *Not*I fingerprinted BAC DNA with probes derived from BAC-end sequences as well as probes corresponding to the *HvHGA4, Hvrh2, Hvpg3, HvHGA1, and HvHGA2* genes (Figure 5-1B; data not shown) showed that the 10 BAC clones form a single physical contig of 350 kb represented by BACs 73D12, 124A1, 14E11, and 68D11 (Figure 5-1B). Clones 124A1 and 14E11, which span the *Rph7* region between the flanking markers Hvgad1 and Hv283 (Figure 5-1A), were chosen for sequencing. Both BACs were sequenced to a coverage of 11X, which resulted in seven contigs for BAC 124A1 and four contigs for BAC 14E11. The gaps were closed by PCR with primers designed at the ends of the contigs. This resulted in two contiguous sequences of 12,258 bp (AY642925) and 184,425 bp (AY642926) separated by a region consisting of LTR retrotransposons of the BARE-1 family. This region could not be assembled to completion because of very high sequence identity between the BARE-1 sequences.

However, some single nucleotide differences and different target site duplications (TSDs) suggested that at least four different BARE-1 sequences are present in this interval, which is ~30 kb long. The 12,258 and 184,425 bp sequences were completely annotated (AY642925-26) using a combination of analytical tools and BLAST searches against the databases.

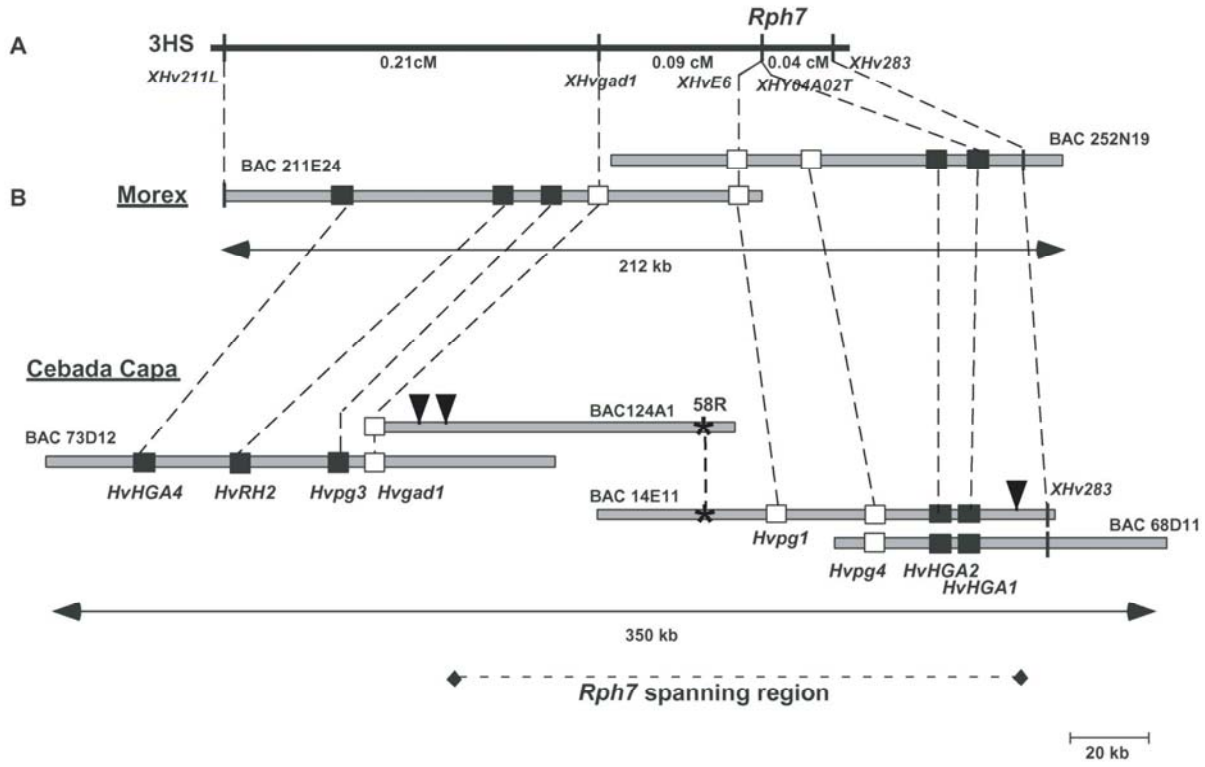


Figure 5-1 Establishment of a 350 kb physical contig at the *Rph7* locus in the resistant cultivar Cebada Capa.

(A) Genetic map of the *Rph7* locus on chromosome 3HS (Brunner *et al.*, 2003). (B) Comparison of the two physical maps in the susceptible cultivar Morex (Brunner *et al.*, 2003) and the resistant cultivar Cebada Capa. Boxes represent the genes located at the *Rph7* locus. White boxes correspond to the three genes (*Hvgad1, Hvpg1* and *Hvpg4*) that were used for screening the Cebada Capa BAC library. The asterisk represents a BAC end probe which was derived from BAC 58F1 and was used to establish the connection between BACs 124A1 and 14E11. The region spanning the *Rph7* gene between the distal and proximal recombination break points (indicated with black vertical arrowheads) is shown as a dotted line.

4.2 Comparison of homologous contigs in Morex and Cebada Capa reveals dramatic and recent genome rearrangements at the *Rph7* locus

The sequences between *Hvgad1* and *Hv283* (Figure 5-1) were compared in the two cultivars Morex and Cebada Capa. This region represents 126.6 kb of sequence in Morex (Brunner *et al.*, 2003; position 85,000 to 211,664 in AF521177) and ~226 kb in Cebada Capa, indicating a large size (~100 kb) difference between the two regions. Seven conserved regions (CR1-7) ranging from 5 to 12.7 kb were identified by dot plot analysis (Appendix 9-2). In total, 53 kb representing 42 and 23% of the sequence in Morex and Cebada Capa, respectively, is conserved with an average of 96% identity. Five conserved regions (CR1 and CR3-6) correspond to sequences containing the *Hvgad1, Hvpg1, Hvpg4, HvHGA1,* and *HvHGA2* genes, whereas CR2 (8.3 kb) and CR7 (5 kb) are intergenic regions (Appendix 9-2). Coding regions account for 16.6 kb (31%) of the conserved 53 kb stretch, whereas the rest (69%) mainly corresponds to 3' and 5' untranslated regions (UTRs) with an average size of 2.7 kb. Large variations in the size of the intervals separating the seven conserved regions are found between the two sequences. In the CR4-CR5 interval, ~22 kb of additional sequence is found in Morex compared with Cebada Capa. In the remaining intervals, additional sequence (100 kb) is found in the intergenic regions of Cebada Capa (Appendix 9-2 and 9-3). The additional sequences correspond to repetitive elements (40%) and to other, putatively non-repetitive, sequences (60%). Consequently, although there are more repetitive elements in Cebada Capa (one CACTA transposon, six solo-LTRs, and three complete LTR retrotransposons) (Appendix 9-2) compared with Morex (one solo-LTR and two complete LTR retrotransposons), the proportion of repetitive sequence in Cebada Capa (50%) is very similar to the one previously found in Morex (55%; Brunner *et al.*, 2003).

To study the structure and evolution of the non-conserved intergenic regions, the seven conserved regions (CR1-CR7) were assembled into single DNA stretches of 53 kb as if they would represent an ancestral sequence present in both cultivars (Figure 5-2A). The position and type of additional sequences that were identified in Morex and Cebada Capa were then compared (Figure 5-2A). Repetitive elements were found in all intervals except in the region between CR4 and CR5 (Figure 5-2A). Most belong to different families, and those belonging to the same family did not correspond to conserved elements. For example, in the CR1-CR2 interval, a BARE-1 retrotransposon is found at the same position in both sequences. However, the presence of different TSDs demonstrates that they do not correspond to the same element. In both cultivars, different types of transposable elements, partial elements as well as duplicated and inverted sequences were identified in the intergenic region between the *Hvgad1* and *Hvpg1* genes (Figure 5-2A; Appendix 9-3). By contrast, only complete retrotransposons and a solo-LTR were identified in the other intergenic regions. Recent retrotransposition of BARE-1 elements in the *Hvgad1-Hvpg1* interval in both cultivars is suggested by the complete identity (100%) of the LTR of BARE-1_211E24-1 in Morex (Brunner *et al.*, 2003) and the very high similarity observed among the different BARE-1 sequences in the 30 kb region in Cebada Capa. In Cebada Capa, sequence expansion in this interval also resulted from two duplications of ~21 and 9 kb flanking the solo-LTR SLB2 and the CACTA transposon Caspar_124A1-1 (Appendix 9-3). Together, these data suggest that the *Hvgad1-Hvpg1* intergenic region was subjected to more rearrangements than the other regions and that despite the difference in sequence composition, this characteristic property is conserved.

The 7.7 kb region between *Hvpg1* and *Hvpg4* (CR3-CR4) is highly conserved (98% identity). This allowed us to identify the exact insertion positions for the different retroelements and to estimate the sequence and timing of insertions and rearrangements that have shaped this region (Figure 5-2B). A BARE-1 solo-LTR and a Bianca retrotransposon were found in this interval in Morex (Brunner *et al.*, 2003), whereas in Cebada Capa, two full-length LTR retrotransposons (Usier_14E11-1 and BARE-1_124A1-1) and two BARE-1 solo-LTRs (SLB3 and 4) were identified (Figure 5-2A). We have determined the number of substitutions in the LTR sequences of the complete retrotransposons. Assuming that at the time of insertion both LTRs of a retrotransposon are 100% identical and using a mutation rate of 6.5 x 10^{-9} substitutions per synonymous site per year (Gaut *et al.*, 1996; SanMiguel *et al.*, 1998), the approximate age of the complete elements was estimated (Table 5-1). This analysis indicates that Bianca_252N19-1 was inserted less than 1 mya (0.92) in the progenitor of Morex, whereas in Cebada Capa, the oldest element Usier_14E11-1 was inserted ~2 mya. Subsequent insertions and intraelement unequal recombination of three BARE-1 retrotransposons followed the insertion of Usier_14E11-1 in Cebada Capa (Figure 5-2B). Our estimates indicate that the complete BARE-1_124A1-1 element inserted ~1.30 mya consistent with its location within Usier_14E11-1. The presence of the SLB3 and four solo-LTRs in BARE-1_124A1-1 indicates retrotransposition and intraelement recombination within the last million years (Figure 5-2B). We have also estimated the age of the conserved 7.7 kb ancestral intergenic sequence using the same molecular clock. Our estimate indicates a divergence time of 1.15 million years (±0.15 million years; data not shown), which is younger than the insertion time of the Usier_14E11-1 retroelement (2.15 million years, ±0.3 million years). A similar discrepancy was found in the

HvHGA2-HvHGA1 interval (data not shown). These discrepancies between the age of the retroelements and the region in which they are inserted support previous suggestions (SanMiguel *et al.*, 1998; Ma and Bennetzen, 2004; Ma *et al.*, 2004) that LTR retroelements evolve at least two times faster than the genes and the UTR regions. If we divide the insertion time estimates of all retroelements in this region by a factor 2, our data suggest high retroelement activity within the last 500,000 years in this region of the barley genome.

No repetitive elements were detected in the CR4-CR5 interval, but a large insertion of 22.2 kb containing the *Hvhel1* gene is found in Morex (Figure 5-2A). This demonstrates that differences between the two barley sequences are not restricted to the repetitive sequence blocks but are also found in the gene composition. In the homologous interval in Cebada Capa, a small sequence of 804 bp corresponding to a duplication of 519 bp of the 3' end and 3' UTR of *HvHGA2* and 285 bp of additional sequence was found (Figure 5-2A). In Morex, the *HvHGA1* and *HvHGA2* genes are separated by an intergenic region of 6.4 kb, whereof 5.6 kb is conserved in Cebada Capa (with 95% identity). An insertion of 15.4 kb that mostly (13.6 kb) corresponds to a complete LTR retroelement (Jelly_14E11-1) interrupted by a Usier solo-LTR (SLU1) occurred in Cebada Capa (Figure 5-2A). Jelly_14E11-1 was estimated to have inserted 1.7 mya, confirming that at the *Rph7* locus, the retroelements are not older than 2 million years. Finally, in the CR6-CR7 interval, 29 kb of additional sequence is found in Cebada Capa compared with Morex. Sequence analysis did not clearly identify repetitive patterns or elements, but short homologies by BLASTX with the hypothetical protein pTREP15 and putative rice transposases or reverse transcriptase indicates the presence of an unidentified element or remaining traces of an older element. In the non-coding CR7 region, (TA) microsatellites and

XI_AF521177-1 elements are conserved at slightly different positions relative to each other in the two regions. Together, our data show a lack of conservation in the length and composition of the intergenic regions between the two barley *Rph7* homologous regions, indicating rapid and recent sequence divergence in the barley genome. Moreover, the fact that no additional genes were identified in the sequence of the resistant cultivar Cebada Capa indicates that *Rph7* is likely a new type of disease *R* gene.

Table 5-1 Estimated times of retrotransposon insertion at the *Rph7* locus.

Elements	K (±SD) [(a)]	Time mya (±SD)
BARE1_211E24-1 (Morex)	0	0
Bianca_252N19-1 (Morex)	0.012 (0.008)	0.92 (0.6)
BARE1_14E11-1 (Cebada Capa)	0.017 (0.003)	1.30 (0.23)
Usier_14E11-1 (Cebada Capa)	0.028 (0.004)	2.15 (0.30)
Jelly_14E11-1 (Cebada Capa)	0.022 (0.006)	1.69 (0.46)

(a) K, estimated number of substitutions per nucleotide site and its standard deviation. K is based on the γ-K2P model for LTR (Kimura, 1980). The average substitution rate of 6.5×10^{-9} substitutions per synonymous site per year that was estimated for the *adh1* and *adh2* genes in maize (Gaut *et al.*, 1996) was applied to estimate the divergence time.

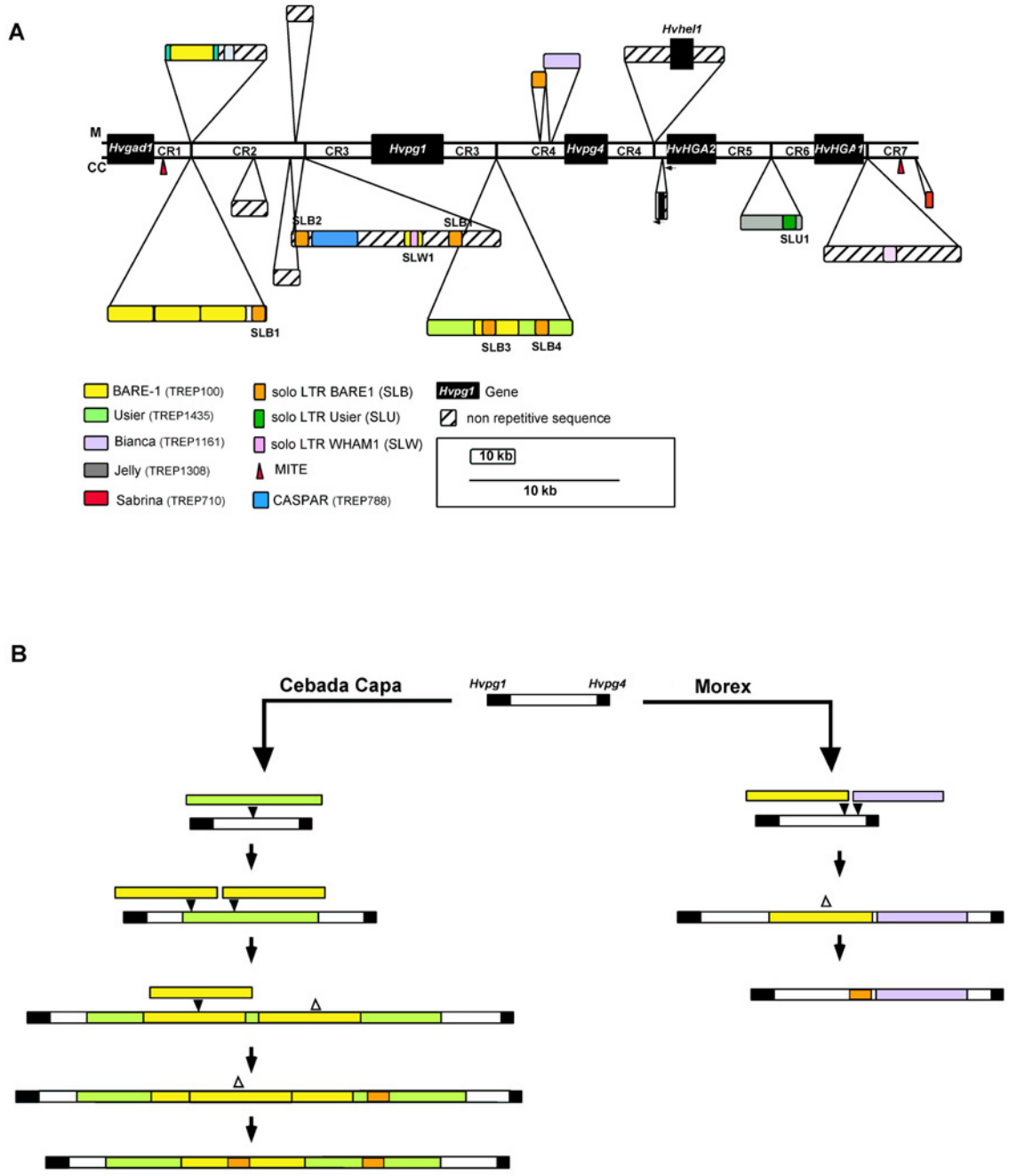


Figure 5-2 Sequence rearrangements in the orthologous intergenic regions in Morex and Cebada Capa.

(A) Schematic representation of 53 kb of DNA sequence corresponding to the assembly of the seven conserved regions (CR1-7) in Morex and Cebada Capa. The upper line (M) represents the 53 kb of sequence from Morex and the lower one (CC) the sequence from Cebada Capa. Filled boxes with color codes given in the figure represent transposable elements which have been inserted in each sequence. Hatched boxes represent non-repetitive sequences which

To examine the effect of such a high level of sequence divergence on recombination in the *Rph7* region, we have analyzed the recombination breakpoints in Cebada Capa and Morex with the assumption that the sequence in Morex is highly similar to the one of Bowman, the parent used for genetic linkage analysis (Brunner *et al.*, 2003). Using a combination of restriction fragment length polymorphism hybridizations and restriction fragment analysis on the BAC sequences, we have identified one recombination break point in a 1.8 kb sequence next to XHv283 in the conserved region CR7 and two recombination events in the CR1-CR2 interval (Figure 5-1). In this interval, ~7.5 kb of the CR1 and CR2 regions as well as sequences from the BARE-1 retroelements are available for homologous recombination. These data indicate that despite overall low sequence conservation between the two haplotypes, there are enough conserved regions that can serve as target sequences for homologous recombination and that recombination is not completely blocked at this locus.

Legend to Figure 5-2 continued:

also differ from the common sequence stretch. Red triangles represent MITES. Genes are represented as black boxes and identified by name. The size scale for the conservedregions (CR1-CR7) is indicated as a line and is three times larger than the one (indicated as a box) used for the non-conserved repetitive and non-repetitive sequences. Two small arrows indicate the duplicated region of 804 bp found near the 3' end of *HvHGA2* in Cebada Capa. (B) Model for the evolution of the *Hvpg1-Hvpg4* intergenic region (CR3-CR4) in Cebada Capa (left) and Morex (right). Color codes are the same as in (A). Filled arrowheads indicate the insertion of an element whereas empty ones indicate unequal recombination between LTR of the same elements leading to solo-LTRs with identical TSD.

4.3 High variability is found in intergenic regions at the *Rph7* locus in the cultivated barley gene pool

To investigate whether the variability observed in the intergenic regions of Morex and Cebada Capa specifically reflects the genetic distance between the two lines or if there is generally high variability in the cultivated barley gene pool, primers were designed in the CR3, CR4, and CR5 regions (Figure 5-3A) and used to amplify genomic DNA from 39 additional cultivated *H. vulgare* lines. PCR amplification across the CR3-CR4 interval (primers cr3-1/cr4-1) produced a 300 bp fragment in Morex, whereas no amplification product was obtained from Cebada Capa because of the insertion of 25 kb of repetitive sequence (Figure 5-3B). The analysis of the 39 lines showed amplification of the 300 bp fragment in 22 lines and no amplification in 17 lines (Table 5-2), indicating the presence of at least two haplotypes represented by Morex (A1) and Cebada Capa (A2) at this locus in the cultivated barley gene pool. A similar strategy was used for amplification in the CR4-CR5 interval. In this case, two sense primers were designed at the 5' end of the *Hvpg4* gene (cr4-2) and on the left border of the insertion point for the 22 kb sequence in Morex (cr4-3), and two reverse primers were designed on the right side of the insertion point of the Morex 22 kb sequence (cr5-1) and on the 3' end of *HvHGA2* (cr5-2) (Figure 5-3A). In Morex, none of the primer combinations resulted in a PCR product because of the 22 kb insertion in this interval. In Cebada Capa, amplification with cr4-3/cr5-1 resulted in a 340 bp fragment, PCR with cr4-2/cr5-1 gave a 4 kb fragment, and the cr4-3/cr5-2 combination lead to the amplification of a 1.6 kb fragment (Figure 5-3B). Amplification with these three different primer combinations resulted in four different types of products in the 39 lines (Figure 5-3B). Eight lines, including near-isogenic lines carrying different *Rph* genes in the background of Bowman, gave the same pattern

as in Morex *(i.e.* no product, B1 haplotype) (Table 5-2). Sixteen lines produced the same fragments as in Cebada Capa (B2 haplotype) (Table 5-2). Among these lines, 14 that originate from Ethiopia, Eritrea, Uruguay, and Argentina carry the *Rph7* resistance allele. Two other lines (Quinn and Magnif104) showed a 60 bp size difference with all primer combinations, indicating a small insertion between primers cr4-3 and cr5-1 (Table 5-2, B3 haplotype). Finally, 13 lines (B4 haplotype) amplified a 0.8 kb fragment with the primers cr4-3/ cr5-2 instead of the expected 1.6 kb fragment, whereas the other fragments had the same size as in Cebada Capa. Cloning and sequencing of the 0.8 kb fragment demonstrated that the size difference results from the absence of the duplicated 804 bp found at the 3' end of the *HvHGA2* gene in Cebada Capa. In summary, at least two haplotypes (A1 and A2) can be found for the CR3-CR4 interval and four (B1, B2, B3, and B4) for the CR4- CR5 interval (Figure 5-3B). Analysis of the haplotype combinations for both intervals in the 41 lines indicated that six (A1B1, A1B2, A1B3, A1B4, A2B2, and A2B4) out of the eight possible combinations were found in the cultivated barley gene pool (Figure 5-3B, Table 5-2). Interestingly, all the lines containing the *Rph7* resistance allele had the same haplotype (A2B2, Table 5-2). The line Sudan also had the A2B2 haplotype, but it is not known to carry the *Rph7* resistance allele. In artificial leaf rust infection tests, Sudan was resistant to the *AvrRph7* isolate, although the resistance phenotype was different than that observed with Cebada Capa and the 14 lines with the A2B2 haplotype (data not shown). These results suggest that Sudan may carry a different resistance allele of *Rph7* and that all lines carrying *Rph7* likely originate from a single donor line.

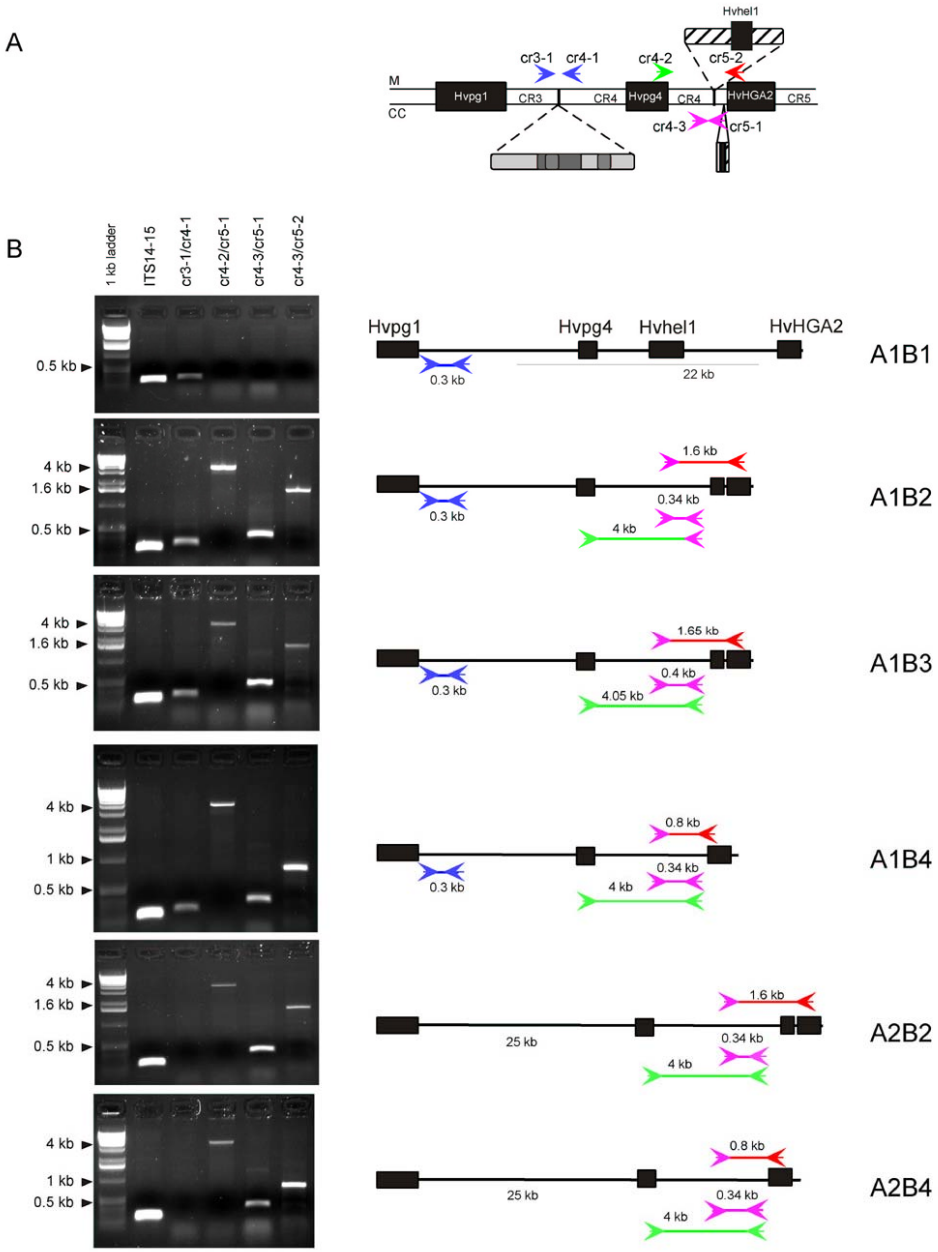


Figure 5-3 Haplotype combinations found in the *Hvpg1-Hvpg4-HvHGA2* intergenic regions in 41 cultivated barley lines.

(A) Schematic representation of the *Hvpg1-Hvpg4* and *Hvpg4-HvHGA2* intervals in Morex (M) and Cebada Capa (CC) showing the position, orientation, and names of the primers used in the analysis. (B) PCR products amplified by four primer combinations and schematic representation of the six haplotype combinations identified in the 41 barley lines. The ITS14-15 primer combination was used as a positive control for PCR amplification. PCR products are represented as lines between the primers with their size above it.

To test whether the variability observed at the *Rph7* locus does not specifically reflect high variability at a fungal disease resistance locus, we have scanned all barley BAC sequences that do not originate from a disease resistance locus and are available from the databases for intergenic regions amenable to PCR amplification. Six intergenic regions that can result in PCR products ranging from 0.5 to 3.9 kb in Morex were identified from BAC sequences originating from the Hordein (AY268139), *Waxy1* (*WX1*) (AF474373), *Vrn1* (Wg644) (AY013246), and *Vrn2* (AY485643) loci. Six primer pairs were designed based on the open reading frames predicted in the annotations and BLASTN analysis against the EST database. In addition, one primer pair was designed in an intergenic region of 1.4 kb originating from a BAC located in the telomeric region of chromosome 3HL between markers MWG838 and MWG010 (N. Stein, personal communication). Out of six primer combinations that amplified the expected product from Morex, three showed the presence of at least two haplotypes in the corresponding intergenic regions in the set of 41 cultivated barley lines. Two to three different haplotypes were found at the orthologous wheat *Vrn2* and *Vrn1* loci on chromosome 4H and 5H, respectively, and two were found at the MWG838-MWG010 locus on 3HL (Table 5-2; Appendix 9-4). The cultivars that belong to a particular haplotype were different at each independent locus (Table 5-2). From these data, we conclude that the variability found at the *Rph7* locus is not specific for this locus but is indicative for a large variability in intergenic regions of the genome from cultivated barley.

Table 5-2 Haplotype combinations in the CR3-CR4 and CR4-CR5 intervals at the *Rph7* locus and haplotypes at the *Vrn2, Vrn1*, and MWG838-MWG010 loci in 41 cultivated barley lines.

		Rph7				
Cultivars	*R* Gene	CR3-CR4	CR4-CR5	*Vrn2*	*Vrn1*	MWG838- MWG010
Morex		A1	B1	C1	D1	E1
Bowman		A1	B1	C2	D1	E1
Bowman*7/Hor2596	*Rph9*	A1	B1	C2	D1	E1
Clipper BC8/5*Bowman	*Rph10*	A1	B1	C2	D1	E1
Bowman*5/Clipper BC67	*Rph11*	A1	B1	C2	D1	E1
Bowman*6/Pl531849	*Rph13*	A1	B1	C2	D1	E2
Bowman*4/Pl584760	*Rph14*	A1	B1	C2	D1	E1
Sundance		A1	B1	C2	D1	E1
Franka		A1	B1	C1	D2	E2
Tunisian		A1	B2	C2	D1	E1
Egypt 4	*Rph8*	A1	B2	C2	D1	E1
Quinn	*Rph2, 5*	A1	B3	C2	D3	E1
Magnif104	*Rph5*	A1	B3	C2	D1	E1
Ribari	*Rph3*	A1	B4	C2	D1	E1
Gold	*Rph4*	A1	B4	C2	D1	E1
Cl1243	*Rph9*	A1	B4	C2	D1	E2
Triumph	*Rph12*	A1	B4	C1	D1	E1
Elisa		A1	B4	C2	D1	E1
JCJ-188		A1	B4	C1	D1	E1
L94		A1	B4	C2	D1	E2
Meltan		A1	B4	C1	D1	E1
Michka		A1	B4	C1	D1	E1
Pallas		A1	B4	C2	D1	E1
Cebada Capa	*Rph7*	A2	B2	C2	D1	E1
Ab1122	*Rph7*	A2	B2	C2	D1	E1
Cebada Forrajera	*Rph7*	A2	B2	C2	D1	E1
Dabat	*Rph7*	A2	B2	C2	D1	E1
Debra Sina	*Rph7*	A2	B2	C2	D1	E1
A2210	*Rph7*	A2	B2	C2	D1	E1
A2211	*Rph7*	A2	B2	C2	D1	E1
A2212	*Rph7*	A2	B2	C2	D1	E1
La Estanzuela	*Rph7*	A2	B2	C2	D1	E1
Hor4445	*Rph7*	A2	B2	C2	D1	E1
Hanka	*Rph7*	A2	B2	C1	D1	E1
Heris	*Rph7*	A2	B2	C2	D1	E1
Ellinor	*Rph7*	A2	B2	C1	D1	E1
Bowman*8/Cebada Capa	*Rph7*	A2	B2	C2	D1	E1
Sudan	*Rph1*	A2	B2	C2	D1	E1
Peruvian	*Rph2*	A2	B4	C2	D1	E1
Bolivia	*Rph2,6*	A2	B4	C2	D1	E1
Lechtaler	*Rph4*	A2	B4	C2	D2	E1

The presence of known leaf rust disease resistance genes in the cultivars is indicated in the second column. Cultivars belonging to one of the six haplotype combinations (AxBx) found at the *Rph7* locus are separated from each other by single lines.

4.4 Evolution of the gene composition at *Rph7* orthologous loci in wheat, barley, rice, and sorghum

The structure and transcriptional orientation of the genes identified in Morex and Cebada Capa are highly conserved. The *Hvpg4, Hvgad1, HvHGA2,* and *HvHGA1* alleles share nucleotide sequence identity ranging from 98 to 99.6% over the entire sequence. The *Hvpg1* alleles are also highly conserved (98% identity at the nucleotide sequence level) except for a deletion of 198 bp at the C terminus of the Cebada Capa allele, which results in the lack of 66 amino acids corresponding to two terminal repeats of a Gly-rich repeat domain. We had previously shown that microcolinearity between barley and rice at the *Rph7* locus was restricted to the conservation of members of the *HGA* gene family on rice chromosome 1 and that genes orthologous to *Hvpg1, Hvpg3, Hvpg4,* and *Hvgad1* are located on the homeologous group 3 chromosomes in wheat (Brunner *et al.*, 2003). A BLASTN search with all the gene sequences against the recently created database of mapped wheat EST [http://wheat.pw.usda.gov/wEST/blast/] allowed us to identify the location of these orthologs in deletion bins on the wheat chromosomes. *HvHGA1* (E value = 1e-126), *HvHGA2* (E value = 1e-75), *Hvpg1* (E value = 1e-79), *Hvpg3* (E value = 3e-22), and *Hvpg4* (E value = 4e-50) orthologs were identified in the three most distal bins of chromosomes 3A, 3B, and 3D (3AS4-0.45-1.00, 3BS8-0.78-1.00 and 3BS1-0.33-0.57, and 3DS6-0.55-1.00), confirming the conservation of the *Rph7* locus composition in barley and wheat (Figure 5-4). In addition, BLAST hits for *Hvpg4* (E value = 4e-50) and *Hvgad1* (E value = 3e-27) were found in bins of chromosomes 4AS, 4BL, and 4DL.

In rice, previous BLASTN and TBLASTX searches had indicated the presence of homologs for *Hvrh2, Hvpg1, Hvhel1, Hvpg3, Hvpg4,* and *Hvgad1* in the rice

genome, but the chromosomal location could not be assessed at that time (Brunner *et al.*, 2003). The recent release of 12 rice pseudomolecules by TIGR [http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml] allowed us to identify the chromosomal location of genes homologous to *Hvrh2, Hvpg3, Hgad1, Hvpg1*, and *Hvpg4* in rice. Using BLASTN, single hits were found for a *pg1* gene (E value = 6e-48) on chromosome 11 and for a *pg3* gene (E value = 2e-18) and two paralogs of *rh2* (E value = 3e-28) on chromosome 2 (Figure 5-4). Four hits were identified with the *Hvgad1* sequence: one on chromosome 3, 46 kb away from a homolog of *Hvpg4* (E value = 3e-49), two for homologous genes located 13.5 kb from each other on chromosome 4, and the best hit (E value = 6e-94) was found with a *gad* gene (*gad8*) on chromosome 8 (Figure 5-4). Thus, these data indicate that putative orthologs of the genes found at the *Rph7* locus in barley are present in rice but are found on five non-colinear chromosomes.

To investigate the origin of this large rearrangement, we have identified homologs for several of these genes and determined their chromosomal location in sorghum, a member of the Andropogoneae subfamily. Sorghum ESTs were identified for the *HvHGA1, HvHGA2, Hvpg1, Hvpg4,* and *Hvgad1* genes and were used to screen six dimensional BAC DNA pools (Klein *et al.*, 2003). The presence of sorghum genes homologous to the barley genes within each positive BAC was confirmed after hybridization with probes for the different barley genes (data not shown). In total, seven BACs were identified. Three (61E22, 63A11, and 52N1) contain a homolog of *Hvpg1*, two (76H6a and 67F2) carry homologous genes to *HvHGA2* and *HvHGA1,* and single BACs have been identified with *Hvpg4* (73H13) and *Hvgad1* (62H18). The chromosomal location of each BAC was determined by integrating the BACs with the sorghum genetic map (Klein *et al.*, 2000) or by BAC-

based fluorescence in situ hybridization (FISH) analysis (data not shown). BACs 76H6 and 67F2 harboring the *HGA1* and *HGA2* genes were located on sorghum chromosome 3 (53 to 55 centimorgan [cM]) in a position colinear to rice chromosome 1 and to the Triticeae chromosome 3 (Figure 5-4). This indicates the conservation of an *HGA* gene family at collinear positions in rice, barley, wheat, and sorghum (group A, Figure 5-4). BACs 73H13 (*pg4*) and 62H18 (*gad*) mapped on sorghum chromosome 1 (190.6 to 193.3 cM) at a position orthologous to rice chromosome 3 (36.1 cM). Homologs of the *pg4* and *gad* genes were also found in bins of chromosome group 4 in wheat (see above). Thus, these two genes are found in colinear regions (group B, Figure 5-4) on wheat chromosome group 4, rice chromosome 3, and sorghum chromosome 1. Finally, the three BACs carrying *pg1* homologs mapped to sorghum chromosome 5 (Menz *et al.*, 2002; Kim *et al.*, 2005), which is syntenic to rice chromosome 11 where the rice *pg1* ortholog was identified (group C, Figure 5-4). Together, these data indicate that the genes in sorghum are not found at a single locus colinear with the Triticeae group 3 but are located on different chromosomes at syntenic positions with rice. The fact that a similar gene distribution is found in members of two different subfamilies (Andropogoneae and Erhartoideae) suggests gene translocation events in the ancestral group 3 chromosomes during the evolution of the Triticeae genomes.

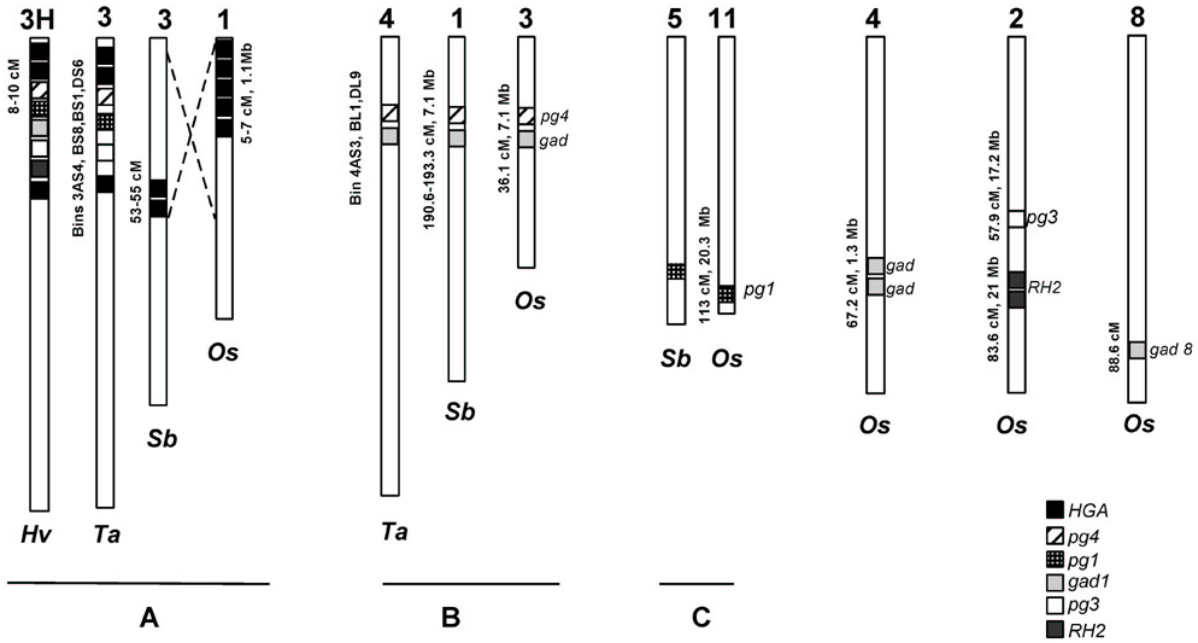


Figure 5-4 Orthologous relationships between the genes located at the *Rph7* locus in barley, wheat, rice, and sorghum.

The location of the genes on genetic maps and/or on physical maps is indicated on the left side of the chromosomes. Hv: *Hordeum vulgare*, Ta: *Triticum aestivum*, Sb: *Sorghum bicolor*, Os: *Oryza sativa*. The inversion identified by Klein *et al.* (2003) between rice chromosome 1 and sorghum chromosome 3 is indicated with interrupted lines. A, B, and C indicate groups of synteny between at least two of the grass species investigated.

5. Discussion

5.1 Intraspecific comparison of homologous loci reveals mechanisms underlying rapid genome evolution in barley

In this study, we have compared more than 300 kb of homologous sequence from two cultivated barley lines that differ in their resistance to leaf rust. Conservation was limited to five genic and two intergenic regions leaving up to 77% divergent sequence between the two loci. This high level of sequence divergence reflects a completely different composition within the intervals spacing the seven conserved regions. In these intervals, neither the type of repetitive elements nor their insertion positions were conserved between the two cultivars. Moreover, one putative helicase gene was absent in Cebada Capa compared with Morex. The difference in the gene composition was not as dramatic as the one recently described in comparative studies between maize inbreds. At the bronze *bz* locus, four genes were missing in one inbred line compared with the other (Fu and Dooner, 2002), whereas at the zein *z1C-1* locus, 10 genes were affected by segmental duplications, insertions, and deletions that have occurred differentially in the two inbreds (Song and Messing, 2003). Additional comparisons at other loci in maize and barley are needed to determine whether the apparent higher level of gene rearrangement observed in maize compared with barley is genome specific as suggested by Song and Messing (2003) or rather locus specific.

Except for a small DNA stretch in the 7.7 kb *Hvpg1-Hvpg4* intergenic region, which was previously annotated as a partial element (George_252N19-1) in Morex (Brunner *et al.*, 2003), there was no evidence for conservation of any repetitive element between the two barley sequences. Although we cannot exclude that part of the conserved regions correspond to new, not yet identified repetitive elements, this

absence of conservation is striking. It particularly contrasts with recent findings in comparative analysis between orthologous loci on wheat homoeologous genomes. At the low molecular weight *GluA-3* locus, a single solo-LTR of the retrotransposon Wilma was conserved between the A and A^m genomes (Wicker *et al.*, 2003), and at the high molecular weight *Glu-1* loci on the A, B, and D genomes, partial sequences of Wilma and Sabrina LTR retrotransposons as well as a stowaway miniature inverted-repeat element were colinear (Gu *et al.*, 2004; Kong *et al.*, 2004). Similarly, a CACTA transposon, an LTR retrotransposon, and five fold-back elements were found at conserved positions on the A and A^m genomes at the *Lr10* orthologous resistant loci (Isidore *et al.*, 2005). Several lines of evidence further support the hypothesis of a very dynamic barley genome with recent activity of transposable elements: (1) estimates of divergence time indicate retroelement movements not older than 1 million years, (2) the majority of the retroelements are complete, and one BARE-1 element has identical LTR, and (3) none of the repetitive elements are conserved. Evidence for an active barley genome was also found by Kalendar *et al.* (2000) who showed that BARE-1 retrotransposition has been induced in response to environmental stress during evolution of wild barley populations in Israel. Halterman and Wise (2004) have also recently shown that members of the *Mla* powdery mildew resistance gene family have been subjected to differential insertion of repetitive elements within and flanking the open reading frames in four barley cultivars. In the maize lineage, it is estimated that retrotransposon invasion occurred in the past 3 million years (SanMiguel *et al.*, 1998; Gaut *et al.*, 2000). Fu and Dooner (2002) have suggested that this invasion has occurred separately in different individuals of the population from which modern maize eventually evolved, resulting in large variability between allelic regions in inbreds. It is possible that similar events are responsible for

the absence of conservation between the two barley cultivars. It is not yet known when retroelement invasion occurred in the barley or the wheat lineages and whether this occurred in one wave such as in maize. However, together with the recent comparative studies between wheat homoeologous genomes (Wicker *et al.*, 2003; Gu *et al.*, 2004; Kong *et al.*, 2004) and between *Mla* alleles in barley (Halterman and Wise, 2004), our data suggest retroelement movements in the different lineages of modern wheat and barley in recent evolutionary times (1 to 3 million years). The complete lack of colinearity between repetitive elements found at barley and wheat orthologous loci (Ramakrishna *et al.*, 2002; Gu *et al.*, 2003) and the low amount of conserved elements between wheat homoeologous loci and barley homologous regions suggest a complete turnover of the repetitive sequences in the Triticeae genomes within a period of less than 4 to 5 million years. Wicker *et al.* (2003) have recently suggested less than 3 million years for the complete elimination of elements in the intergenic regions in wheat.

The average divergence time estimates for most of the genes present at the *Rph7* locus (1.2 million years; data not shown) suggest a radiation time for wild barley lineages comparable to the maize lineages (1 to 2 million years) (Gaut and Clegg, 1993) and to the 0.5 to 1 million years estimated for the A genomes of wheat (Huang *et al.*, 2002). This is in the same time frame as the estimated divergence time for the two rice subspecies *O. sativa* ssp *indica* and *japonica* (Bennetzen, 2000; Ma and Bennetzen, 2004). However, in contrast with the high level of variability observed between the progenitors of modern maize and barley cultivars, high conservation has been found between orthologous sequences in the two rice subspecies (Song *et al.*, 2002; Han and Xue, 2003; Ma and Bennetzen, 2004). Thus, these data indicate large differences in the activity of the different grass genomes in the last 2 million years of

evolution and support the idea that rice has a more stable genome than other grasses (Ilic *et al.*, 2003). Finally, our data and the recent findings in maize (Fu and Dooner, 2002; Song and Messing, 2003) also demonstrate that intraspecific comparisons can be as informative as interspecific comparisons in revealing the mosaic organization of orthologous sequences in grass genomes.

Despite high sequence divergence, interesting features were conserved in intergenic regions of both cultivars. First, in the two sequences, some regions (*e.g. Hvpg4-HvHGA2* interval) seem to be protected from transposable element invasion, whereas others, such as the *Hvgad1-Hvpg1* intergenic regions, have complex patterns of sequence rearrangements and contain several partial sequences of repetitive elements. Mechanisms underlying the preferential insertion of repetitive elements at target sites in plant genomes are not yet well understood, and except for repeat sequences of knob DNA in maize (Ananiev *et al.*, 1998), no particular sequences have been specifically associated yet with the insertion of repetitive elements. In this respect, the conservation of the insertion position for the BARE-1 retrotransposons in the CR1-CR2 interval is very interesting. Thus, intraspecific comparisons may be very helpful in identifying conserved regions with preferential insertion of transposable elements. A second interesting trend is the conservation of a ratio of ~50% in the proportion of repetitive to non-repetitive sequence in both regions. This is lower than the average 70% found at other loci in barley (Dubcovsky *et al.*, 2001; Rostoks *et al.*, 2002) and might therefore be locus specific, reflecting particular evolutionary constraints associated with the presence of essential genes in this gene-rich region.

Finally, this comparison has allowed us to determine that there are no additional candidate genes for *Rph7* at the resistance locus in Cebada Capa, demonstrating

that *Rph7* belongs to a new type of disease resistance gene. Complementation experiments to identify *Rph7* among the four candidate genes *Hvpg1*, *Hvpg4, HvHGA1,* and *HvHGA2* are currently underway.

5.2 Large haplotype variability in the cultivated barley gene pool

Our analysis indicates high diversity at the *Rph7* locus in the cultivated gene pool. The relative representation of the different haplotype combinations in 41 *H. vulgare* lines shows that the combinations observed in Morex and in Cebada Capa do not correspond to the predominant (A1B4) haplotype. In fact, the A1B1 combination of Morex was only found in three other lines, and the A2B2 combination of Cebada Capa was detected in the lines known to carry the *Rph7* resistance allele. The presence of a single resistant haplotype for *Rph7* in cultivars that originate from Ethiopia, Eritrea, Uruguay, and Argentina suggests a single origin of the *Rph7* gene, very likely in the Fertile Crescent, the center of origin of barley (Badr *et al.*, 2000). We are currently analyzing a larger set of 200 wild barley lines from Israel and the Fertile Crescent to determine to what extent the diversity observed in the cultivated gene pool originates from the wild relatives and to study the relative abundance of the *Rph7* resistance haplotype in the wild gene pool to understand the origin of this important disease resistance gene.

So far, genetic diversity studies in cultivated barley have been performed with different DNA marker techniques and have given contradictory results depending on the locus or the marker type used (Graner *et al.*, 2003). Here, we show that PCR-based analysis of intergenic regions can also be used to detect the genetic diversity present in the cultivated barley germplasm. Our results support and extend recent

results of Halterman and Wise (2004), indicating variability in flanking regions of the *Mla* genes. Moreover, we show that variability in intergenic regions is not restricted to disease resistance loci, which are often considered as more variable than other loci, but is also found at three independent loci. Further BAC sequence comparisons between barley homologous regions should help to confirm these data and to determine whether variability is mainly based on transposable element activity in the intergenic regions or also involves gene loss and gene relocation, such as in maize, with a possible impact on gene dosage and recombination distribution along the chromosomes. Furthermore, expression studies of the genes found at the *Rph7* locus in Morex and Cebada Capa will be performed to determine whether, similar to the finding of Song and Messing (2003) in maize, allelic gene expression is differentially regulated in non-colinear barley haplotypes.

5.3 A complex history of rearrangements involving gene movements is responsible for the gene diversity at the *Rph7* locus in the Triticeae

The *Rph7* locus is a diverse gene-rich locus and represents a very good model for studying complex gene rearrangements during the evolution of grass genomes. Our comparative analysis in barley, wheat, sorghum, and rice suggests that the gene composition observed in barley and wheat results from gene movements specifically in the Triticeae lineage. The orthologous relationships found in sorghum and rice indicate that the gene distribution in these species likely reflects an ancestral pattern. As the number of comparative studies between rice and other grass genomes increases, there are more examples of genes that are conserved in the different genomes but are found at non-orthologous positions on the chromosomes (Li and

Gill, 2002; Song *et al.*, 2002; Ilic *et al.*, 2003). Recently, Guyot *et al.* (2004) have analyzed the conservation of more than 200 ESTs mapped in deletion bins on the short arm of wheat chromosome 1AS with the sequence of the 12 rice pseudomolecules. Less than 20% of the ESTs were found in colinear regions on rice chromosome 5S, whereas the remaining ESTs were distributed at non-orthologous loci on all other rice chromosomes. Song *et al.* (2002) have suggested a possible mechanism underlying gene movements based on gene amplification followed by illegitimate recombination. Interestingly, the *gad* and *rh2* genes, which have been relocated on chromosome 3HS at the *Rph7* locus in barley, belong to gene families in barley and rice and are duplicated on rice chromosomes 4 and 2.

Several BAC libraries are now available from wheat, and a BAC library from *Brachypodium sylvaticum* that represents an interesting intermediate between the rice and Triticeae genomes has been recently constructed (Foote *et al.*, 2004). Further sequence comparisons at the *Rph7* orthologous regions from sorghum, wheat, and Brachypodium should provide relevant information to support or reject the hypothesis of gene movements and to understand molecular mechanisms at the origin of the gene composition at the *Rph7* locus in the Triticeae.

VI. FUNCTIONAL ANALYSIS OF THE CANDIDATE GENES FOR *RPH7*

1. Introduction

The *Rph7* leaf rust resistance gene was genetically mapped within an interval of 0.13 cM on chromosome 3HS in barley (Brunner *et al.*, 2003). Sequencing of a 212 kb physical BAC contig from the susceptible cultivar Morex spanning the genetic interval identified five candidate genes for *Rph7*. None of them showed similarity to known disease resistance (*R*) genes, suggesting that *Rph7* is either a new type of resistance gene or that it is absent in the cv. Morex (Brunner *et al.*, 2003). To test these hypotheses, we have constructed a BAC library of the resistant *Rph7* donor line Cebada Capa (Isidore *et al.*, 2005) and subsequently established and sequenced a BAC contig of 350 kb spanning the resistance locus (see chapter V). Sequence comparison between the homologous *Rph7* regions in Morex and Cebada Capa did not result in the identification of any additional gene in the resistant cultivar. On the contrary, one of the predicted genes in Morex (*Hvhel1*, see chapter V) was absent in Cebada Capa. Thus, four genes, *Hvpg1, Hvpg4, HvHGA1,* and *HvHGA2,* are candidates for *Rph7*. Different strategies can be followed for the identification of a disease *R* gene from a set of candidate genes in plants.

Sequence comparison of resistant and susceptible (natural or mutation induced) alleles has been used to identify several disease resistance genes in plants. For example, the *Rpg1* stem rust resistance gene was identified by sequence comparison of candidate gene alleles from 9 resistant and 19 susceptible barley cultivars (Brueggeman *et al.*, 2002). Sequence comparison between ethyl

methanesulfonate (EMS)-induced mutants and wild type alleles of two candidate candidate genes was also used to support the identification of the wheat leaf rust resistance gene *Lr10* gene (Feuillet *et al.*, 2003).

Differential expression studies can as well potentially help to identify genes among candidates. However, *R* genes are often expressed constitutively in uninfected plants and there are only few examples where *R* gene expression is induced upon infection (Yoshimura *et al.*, 1998; Levy *et al.*, 2004; Radwan *et al.*, 2005). Resistant and susceptible alleles of the *R* gene *Xa27* from rice encode identical proteins (100% identity at the DNA sequence level). However, only the resistance allele is expressed, and expression of *Xa27* is only detected when a rice plant is challenged by bacteria harboring *AvrXa27* (Gu *et al.*, 2005).

Complementation experiments either through transient or stable transformation systems provide the ultimate proof of gene identity. In barley, stable transformation can be performed either through biolistic or via *Agrobacterium tumefaciens*-mediated DNA delivery. Particle bombardment was the first method developed for efficient transformation of barley (Wan and Lemaux, 1994). Later on, stable transformation was also demonstrated by co-cultivation of immature embryos with *Agrobacterium tumefaciens* (Tingay *et al.*, 1997; Patel *et al.*, 2000). Travella *et al.* (2005) have recently compared the two transformation systems. They concluded that the *Agrobacterium*-mediated method offers significant advantages over particle bombardment, with higher transformation efficiency and transgenic plants showing simpler integration patterns, fewer rearrangements, and reduced silencing. *Agrobacterium*-mediated barley transformation was recently used to demonstrate the function of the ‘eukaryotic translation initiation factor 4E’ *(Hv-eIF4E)* in recessive

resistance to *Barley mild mosaic virus* (BaMMV) mediated by the *rym4* gene (Stein *et al.*, 2005).

However, stable transformation is usually a long and laborious process in plants species with long life cycles and is not routinely performed in all plants. Transient assays represent an interesting alternative to assess *R* gene function. Functional assessment of candidate genes for powdery mildew resistance have taken advantage of a transient transformation assay based on the bombardment of leaf epidermal cells with a plasmid carrying the test gene and a second plasmid containing a reporter gene (Schweizer *et al.*, 1999). The identity of the barley powdery mildew resistance genes *Mla6* and *Mla1* was demonstrated using such a single-cell expression assay (Halterman *et al.*, 2001; Zhou *et al.*, 2001). In wheat, Yahiaoui *et al.* (2004) have used a similar assay to demonstrate the identity of the *Pm3b* wheat powdery mildew resistance gene. However, one limitation of this technique is that most of the DNA is delivered into the epidermal cells and very few DNA can reach the mesophyll cells. It is therefore not suitable for assessing the function of genes against pathogens such as leaf rust that develop only in the mesophyll.

Virus-induced gene silencing (VIGS) offers an alternative to these limitations with a reverse genetic approach that allows to study genes through a loss of function. VIGS occurs when plants are infected with a virus carrying target sequences with homology to a host nuclear gene. The virus infection triggers the cytoplasmatic degradation of any RNA with sufficient homology to the target sequence, resulting in post-transcriptional silencing of the homologous nuclear gene(s). This phenomenon is also known as RNA interference (RNAi). *Barley stripe mosaic virus* (BSMV), a single-strand RNA virus with a tripartite genome composed of α, β, and γ RNAs, was

described as a useful vector for VIGS in barley (Holzberg *et al.*, 2002; Lacomme *et al.*, 2003). Scofield *et al.* (2005) have very recently demonstrated BSMV-mediated silencing of the wheat leaf rust resistance gene *Lr21* and of genes functioning in the *Lr21*-mediated resistance pathway in hexaploid wheat.

We have evaluated the four candidate genes for *Rph7* by *Agrobacterium*-mediated transformation of the leaf rust susceptible barley cv. Golden Promise with the alleles of the resistant cv. Cebada Capa. In addition, VIGS was used as an alternative technique to test two of the *Rph7* candidate genes.

2. Materials and Methods

2.1 Plasmid constructs for *Agrobacterium*-mediated barley transformation

Genomic DNA fragments of 5.8 kb and 6.4 kb containing the entire coding region of *Hvpg4* and *HvHGA1*, respectively, and ~2 kb upstream and ~1 kb downstream sequences were amplified from Cebada Capa BAC 14E11 with a *PfuUltra* high-fidelity DNA polymerase (Stratagene Europe, Amsterdam, the Netherlands) according to manufacturer's protocol. *Not*I and *Asc*I restriction sites (underlined) were introduced by PCR with the following primers (numbers in brackets indicate positions on the Cebada Capa BAC sequence AY642926): PG4-Not: 5'-CAAAGCGGCCGCGGGCCGATCCATACACTGGT-3' (115,939-115,958); PG4-Asc: 5'-CTAGGGCGCGCCACCTCCATGGCTAGCAGTCTAGACTT-3' (110,154-110,179); HGA1-Not: 5'-CCAAGCGGCCGCCCACAGTTCTGGTTGCTGTATGGTAT-3' (149,850-149,875); and HGA1-Asc: 5'-AATCGGCGCGCCGTCCCTATTTAAGTCAGCCGAAAC-ATT-3' (143,493-143,519). The PCR products were cloned into the pCR-Blunt II-TOPO Vector (Invitrogen AG, Basel, Switzerland) and sequenced to exclude PCR errors. Inserts were released by *Asc*I/*Not*I digestion and subcloned into the *Agrobacterium* binary vector pWBVec8+A (described by Bieri *et al.*, 2004 and kindly provided by Dr. Stéphane Bieri; Figure 6-1).

A 7.4 kb *Xho*I/*Cla*I genomic fragment originating from Cebada Capa BAC 14E11 and including the coding region of *HvHGA2*, 3.2 kb of 5' upstream, and 1.1 kb of 3' downstream sequences was inserted into a modified pSP72 Vector (Promega, Wallisellen, Switzerland) which contains unique *Asc*I and *Not*I sites flanking the multiple cloning site. The *Asc*I-*Not*I fragment was subsequently transferred into the pWBVec8+A vector (Figure 6-1).

Subcloning of *Hvpg1* was performed in two steps. First, a 2.4 kb *Xba*I/*Sal*I fragment of BAC 14E11 corresponding to the promoter region of *Hvpg1* was subcloned into the modified pBluescript Vector (Stratagene) pBlue+AN, containing flanking *Asc*I/*Not*I sites, to generate the plasmid pBlue-PG1pro. In a second step, a 5.8 kb fragment containing the coding region of *Hvpg1* and 1.3 kb of downstream sequence was amplified from BAC 14E11 with the primers PG1-Sal: 5'-GAAAGAT-GGACGCGGAGCAGGA-3' (79,790-79,811) and PG1-Apa: 5'-CTCAGAGAGATCAA-CTCTGTAATTCTCATCCACAT-3' (73,779-73,813) and inserted into the respective restriction sites of pBlue-PG1pro to generate the plasmid pBlue-PG1comp. The resulting construct was subcloned as *Asc*I-*Not*I-fragment into the Gateway® entry vector pENTR™ 4 (Invitrogen). Gateway® Technology (Invitrogen) was subsequently used to transfer the insert into the binary destination vector pWBVec8+Ga-b (kindly provided by Dr. Stéphane Bieri) according to the manufacturer's protocol (Figure 6-1).

2.2 *Agrobacterium*-mediated barley transformation

Agrobacterium tumefeciens-mediated transformation of immature embryos (~200 embryos per plasmid construct) derived from *Hordeum vulgare* ssp *vulgare* (barley) cultivar Golden Promise was performed as described (Tingay *et al.*, 1997; Wang *et al.*, 2001; S. Schulze and H. Steinbiss, unpublished data).

2.3 Characterization of transgenic plants

Transgenic T_0 and T_1 plants were analyzed by DNA gel blots. Isolation of genomic DNA from young leaf tissue and Southern hybridization were performed as

described by Stein *et al.* (2000). DNA was digested with a restriction enzyme that cuts within the transformed construct and in the adjacent genomic DNA, thereby producing different fragment sizes depending on the place of insertion of the transgene (Figure 6-1). Membranes were hybridized with probes that were derived from restriction digests of the plasmid constructs and are located in between the restriction site and the left or right T-DNA border, thus allowing determination of the copy number of the transgene (Figure 6-1). Artificial infection of T_1 plants with barley leaf rust isolate 1.2.1 avirulent on *Rph7* was performed as described by Brunner *et al.* (2000).

2.4 Analysis of gene expression in transgenic plants

Total RNA was extracted from leaves of T_1, Cebada Capa and Golden Promise plants using the TRIzol reagent (Invitrogen AG, Basel, Switzerland) according to the provider's instructions. Poly A^+ mRNA was isolated with the Oligotex® mRNA kit (Qiagen AG, Basel, Switzerland) and 0.2 µg mRNA was subjected to reverse transcription with an oligo(dT) 21-mer primer. The primer was annealed by incubating the reaction for 5 min at 70°C followed by 2 min on ice. First-strand synthesis was performed in a 30 µl volume after adding 1 x buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 1.5 mM dNTPs, 40 units of RNAse inhibitor (RNaseOUT, Invitrogen), 10 mM DTT, and 200 units of SuperScript™ II Reverse Transcriptase (Invitrogen). After incubation for 1 h 30 min at 42°C, the reaction was stopped at 70°C for 15 min. PCR was performed on 1 µl of the first-strand cDNA product and 1 µl of the first PCR product for nested PCR in 20 µl containing 1 unit of *Taq* DNA-polymerase (Sigma-Aldrich, Buchs, Switzerland), 1 x PCR buffer (10mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.001% gelatin), 100 µM dNTPs, and

500 nM primers described in Table 6-1. Amplifications were performed in a PTC-200 thermocycler (MJ Research, Bioconcept, Switzerland) as follows: 3 min at 95°C, 30 cycles of 45 s at 95°C, and 45 s at 63 to 70°C depending on the primer combination (see Table 6-1) followed by 25 s to 1 min 30 s (see Table 6-1) at 72°C. The extension of the amplified products was achieved at 72°C for 2 min. PCR products were separated by electrophoresis on 2 to 3% agarose gels and visualized under UV light after ethidium bromide staining.

Table 6-1 Primers used for RT-PCR experiments

Gene	Cebada Capa allele-specific				Non allele-specific			
	Primer	Sequence (5'-3')	T_m	Et	Primer	Sequence (5'-3')	T_m	Et
			(a)	**(b)**			**(a)**	**(b)**
Hvpg1	SNP1	GCAGTTGTGGAGTTATACTTACTGG	63	60				
	SNP2	GCTGCTGCAGTATCTTCAGCATGC						
	PG1-H	GTGGCAGCATGGTTATTGA	65	90				
	PG1-O2	TCCCATGATCATGCCACCACATC						
Hvpg4	SNP1	GACGAGTGGGAGGGGAG	63	30	PG4-7	GACCTCGAAGAAGGTGGTGAC	70	25
	SNP3	TGCAGATGGAAGTATAATTCGGGA			PG4-2new	TCGCCTCTTGGTGCGAGTTC		
HvHGA1	SNP1	GGAGGAAGAGACCCCGAAGAGA	65	30				
	SNP2	GGCTGAATCTTCCACTCCTGGCT						
HvHGA2	SNP2b	AACGGCAAGGTGGTTTGCAT	65	45	HGA2-f	CTCAAGAGCAGCGGCAGCGG	65	40
	SNP5	ATCGTCATGGTCCAGGTCCAC			HGA2-g	AGGCCCACGGTGACGCGCC		

(a) Annealing temperature in °C
(b) Extension time in sec.

2.5 Virus-induced gene silencing in barley

Three cDNA fragments for *Hvpg1* (pg1-1 to pg1-3) and one fragment for *HvHGA1* (hga1-1) were amplified from Cebada Capa cDNA using 2 µl of first-strand cDNA product (see 2.4) as a template (Table 6-2). PCR was performed in a 25 µl

volume containing 1 unit of *TaKaRa Ex Taq*™ DNA-polymerase (TaKaRa Bio Inc., Dalian, Japan) 1 x PCR buffer, 250 µM of each dNTP (TaKaRa), and 400 nM primers listed in Table 6-2. Amplifications were done as follows: 3 min at 94°C, 35 cycles of 10 s at 98°C, 30 s at 60°C, and 30 s at 68°C, ending with an extension step of 5 min at 72°C. The amplified PCR products were excised from an agarose gel, digested with *Pac*I/*Not*I and cloned in anti-sense orientation into a *Pac*I/*Not*I digested pBSMV-γb vector. The plasmids used in these experiments are based on the constructs described by Holzberg *et al.* (2002). *In vitro* transcription of viral RNAs and plant inoculation were performed as described by Scofield *et al.* (2005). Artificial leaf rust infection with barley rust isolate 1.2.1 was previously described (Brunner *et al.*, 2000).

Table 6-2 Primers used to amplify cDNA fragments of the Cebada Capa genes for the VIGS experiment

Primer	Sequence (5'-3') [(a)]	Amplified gene fragment	
		VIGS construct	PCR product length
PG1-V1f	TTACATTTAATTAATGATGAGCCTCTGGTTGTTCCGC	pg1-1	274 bp
PG1-V1r	ATATTAGCGGCCGCTGCCTCCTGCGGTACCATCGC		
PG1-V2f	ACTTAATTAATTAATTGTGGACATCAAGAACTTACCATCTGC	pg1-2	339 bp
PG1-V2r	TATTTAGCGGCCGCGGGAACCAGCATAAGCGCCGC		
PG1-V3f	TACCATTTAATTAATCTGCTTCCCTGGATCGCGATG	pg1-3	324 bp
PG1-V3r	TATTTCGCGGCCGCTAGCTCATGGTGACCACAGAGACAC		
HGA1-V1f	AATAATTTAATTAAATCCAGGCTCCCGCCGCC	hga1-1	218 bp
HGA1-V1r	TTAACTGCGGCCGCGGCCCCTCCGTGGTACACACC		

(a) *Pac*I and *Not*I restriction sites, respectively, that were used to clone the amplified fragments into the pBSMV-γb vector are underlined.

2.6 Screening for leaf rust susceptible mutant plants

About 3,000 seeds of a double haploid (DH) line generated from the near-isogenic line (NIL) Bowman8*/3/ND771/Cebada Capa/MT81195 were irradiated with 150 Gy of γ-ray at the International Atomic Energy Agency (IAEA) reactor in Seibersdorf, Austria. Seeds were field-propagated and M2 plants phenotypically analyzed by artificial infection with the barley leaf rust isolate avirulent on *Rph7* as described in Brunner *et al.* (2000).

2.7 Re-evaluation of recombination events with genomic STS markers

Primers amplifying polymorphic fragments in Cebada Capa and Bowman were designed based on the comparison of the homologous Cebada Capa and Morex sequences. PCR was performed on 50 ng of genomic DNA in 20 µl reaction containing 1 unit of *Taq* DNA-polymerase (Sigma-Aldrich, Buchs, Switzerland), 1 x PCR buffer (10mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.001% gelatin), 100 µM dNTPs, and 400 nM primers described in Table 6-3. Amplifications were performed in a PTC-200 thermocycler (MJ Research, Bioconcept, Switzerland) as follows: 3 min at 95°C, 30 cycles of 30 s at 95°C, and 30 s at 59 to 65°C depending on the primer combination (see Table 6-3) followed by 40 s to 1 min 10 s (see Table 6-3) at 72°C. The extension of the amplified products was achieved at 72°C for 2 min. PCR products were separated by electrophoresis on 2% agarose gels.

Table 6-3 Primers used to amplify genomic STS markers for the re-evaluation of the recombination events

Marker	Primer	Sequence (5'-3')	Fragment size in Cebada Capa	Fragment size in Morex (Bowman)	T_m **(a)**	Et **(b)**
LE-1f/1r	pLE-1f	AAGGGCTCTAGGGTCTGTTC	985 bp	843 bp	59	70
	pLE-1r	GATCAGTCTGGGAGTTACCAC				
GAD-1f/2r	pGAD-1f	TGCCGGTGCCAACAACATGT	590 bp	430 bp	60	40
	pGAD-2r	ATGGGCATTCGACTCGATGC				
RE-1f/2r	pRE-1f	CAGTTGCCACGCATGTTC	696 bp	433 bp	60	60
	pRE-2r	CATTAGACATGGGTGCAAGTAC				
RE-3f/3r	pRE-3f	TTCGTTGGGCCTACTCACATTG	808 bp	867 bp	65	60
	pRE-3r	GAAGAATAACTATTCCTCACCCAGG				

(a) Annealing temperature in °C
(b) Extension time in sec.

3. Results and Discussion

3.1 *Agrobacterium*-mediated transformation of the susceptible barley cv. Golden Promise

Four genes, *Hvpg1, Hvpg4, HvHGA1,* and *HvHGA2,* were identified as candidates for *Rph7* in the resistant barley cultivar Cebada Capa (see chapter V). To identify the resistance gene among them, each of the four candidate genes was transformed separately into the leaf rust susceptible barley cv. Golden Promise. For each candidate gene, *Agrobacterium tumefaciens*-mediated transformation was performed with plasmid constructs carrying a genomic fragment containing the entire coding region, ~2 kb upstream and ~1 kb downstream sequences (Figure 6-1). Fifty to 96 T_0 plants (Table 6-4) per construct were obtained. Genomic DNA was extracted and analyzed by Southern blot hybridization to identify the transgenic plants and to determine the copy number of the transgene (Figure 6-2). As gene silencing frequently occurs when several copies of the transgene are present, only plants showing a single insertion event were further analyzed. For each construct, six to 24 progeny plants originating from about ten independent T_0 single copy transformation events were artificially infected with a barley *AvrRph7* leaf rust isolate (Table 6-4).

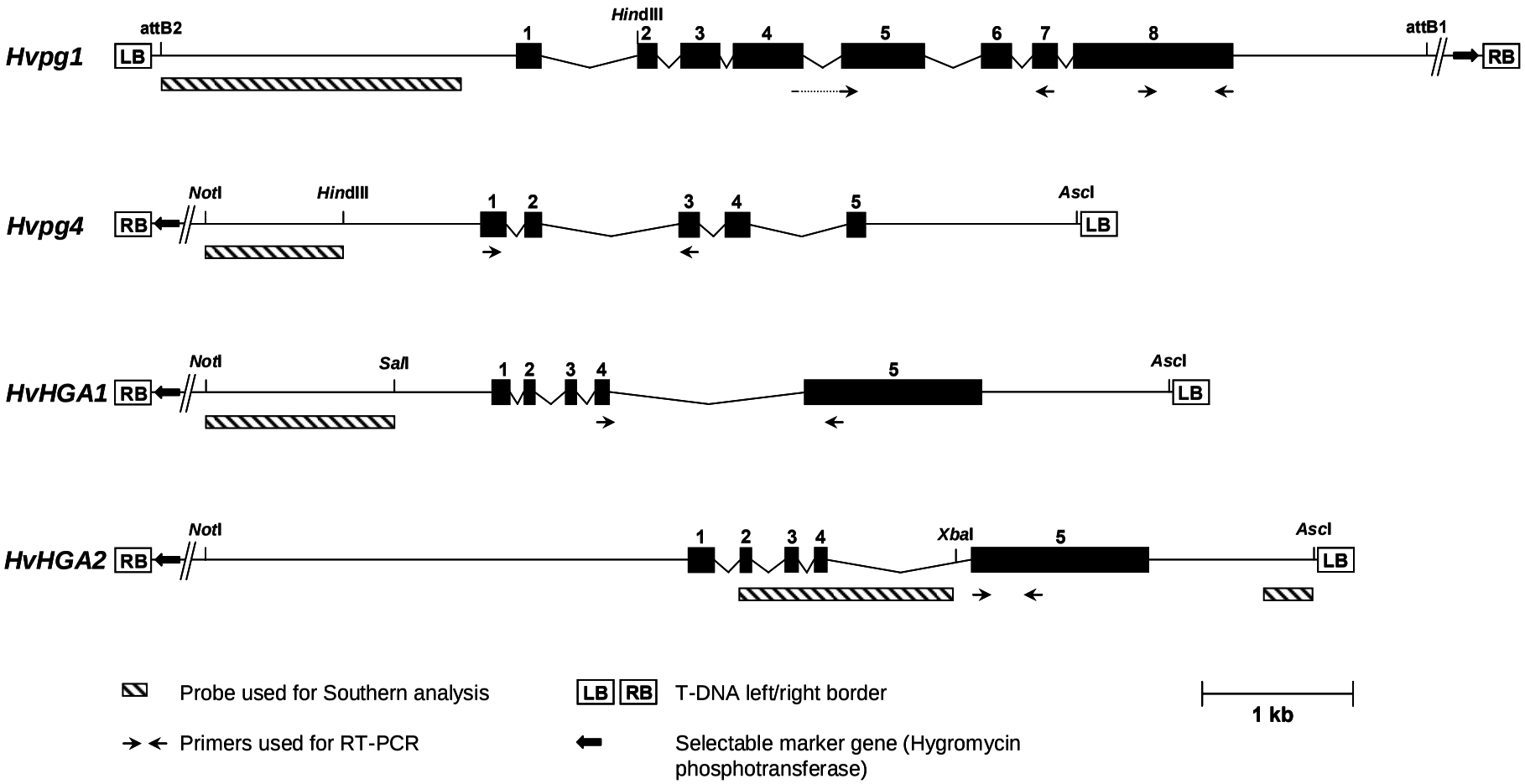


Figure 6-1 Plasmid constructs of the four Cebada Capa candidate genes used for *Agrobacterium*-mediated transformation of barley cv. Golden Promise

Genomic DNA fragments cloned between *Not*I/*Asc*I restriction sites and attB2/attB1 recombination sites, respectively, of the *Agrobacterium* binary vector pWBVec8+A are shown. Black boxes with numbers above them represent exons of the genes. The location of probes and restriction sites used for Southern analysis of the transgenic plants is indicated. Arrows represent primers used for RT-PCR experiments.

Table 6-4 Summary of the transformation experiments with the four candidate genes for *Rph7*.

Transgene	T_0 plants	Confirmed transgenic plants	Transformants with single insertion of the transgene	T_1 families [(a)]	Total T_1 plants [(b)]
Hvpg1	94	61	22	11	154
Hvpg4	71	64	33	7	71
HvHGA1	96	48	13	11	178
HvHGA2	50	37	14	10	117

(a) Total number of T_1 families (each originating from an independent T_0 transformant) that were artificially infected with leaf rust

(b) Total number of pheno- and genotypically analyzed T_1 plants

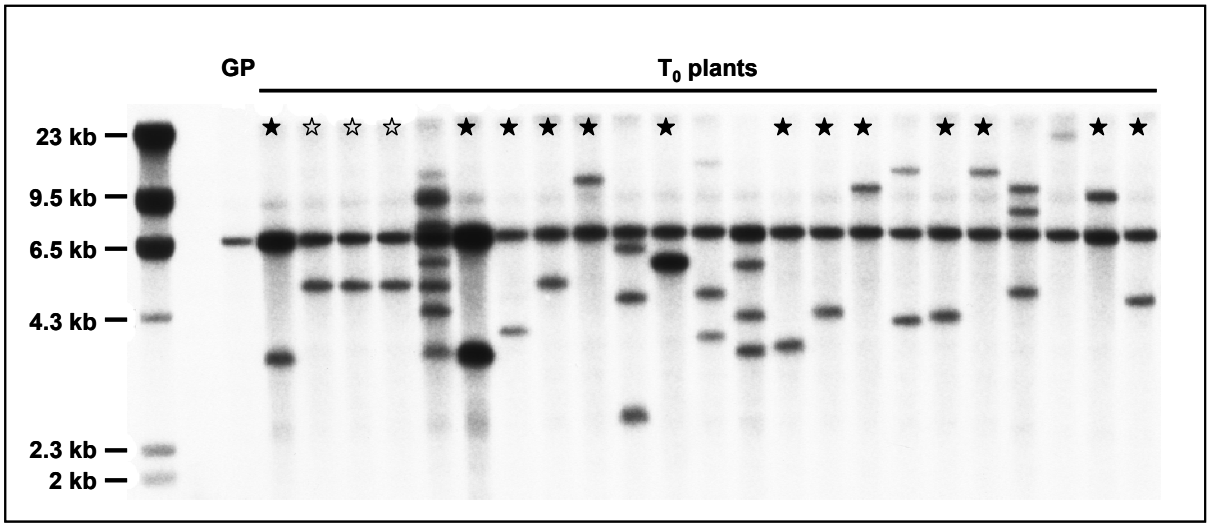


Figure 6-2 Southern blot analysis of barley cv. Golden Promise transgenic T_0 plants

Southern hybridization of 23 T_0 plants obtained from *Agrobacterium*-mediated transformation with the *Hvpg4* construct. The non-transformed Golden Promise control is indicated as GP. Genomic DNA was digested with *Hin*dIII and a 922 bp *Hin*dIII/*Xba*I-fragment of the *Hvpg4* construct was used as probe (Figure 6-1). Filled asterisks indicate individual T_0 plants that show a single insertion of the transgene. Open asterisks indicate plants that derive from the same transformation event (sister plants).

None of the 520 analyzed transgenic T_1 plants showed a resistance phenotype, indicating that none of the four candidate genes for *Rph7* was able to complement the susceptible phenotype of the cultivar Golden Promise. Southern blot analysis revealed that the majority of the T_1 families segregated for the presence of the transgene, demonstrating that the lack of resistance reaction was not due to a loss of the transgenes from the T_0 to the T_1 generation (Figure 6-3).

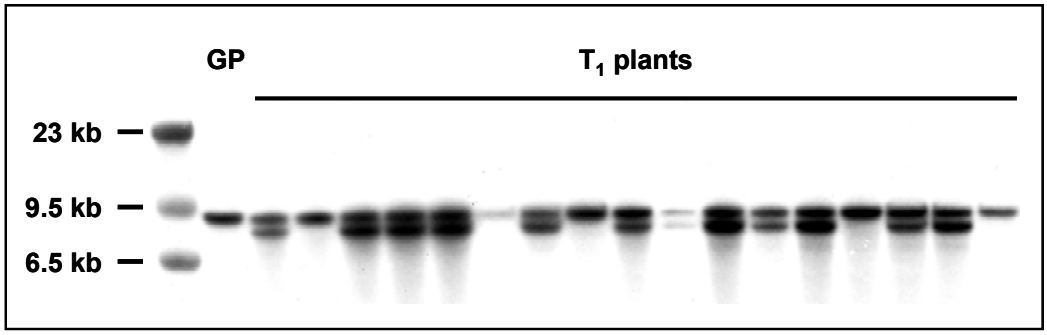


Figure 6-3 DNA hybridization of Golden Promise transgenic T_1 plants

Segregation of the *Hvpg1* transgene in a T_1 family of 17 plants. Genomic DNA was digested with *Hin*dIII and a 2 kb *Eco*RV fragment of the *Hvpg1* construct was used as probe (Figure 6-1). The non-transformed Golden Promise control is indicated as GP.

3.2 Expression analysis of the transgenes by RT-PCR

The expression of the transgenes in the T_1 generation was analyzed by Reverse transcription-PCR (RT-PCR) experiments. For each construct, two to three different T_1 families, originating from individual T_0 transformation events, were analyzed. Messenger RNA (mRNA) was extracted from leaves of two to four plants per T_1 family. Two T_1 plants identified to carry the transgene and one non-transgenic plant as a negative control were selected from each T_1 family for expression analysis. DNA hybridization revealed that alleles of the four candidate genes are also present

in the susceptible cultivar Golden Promise. In order to distinguish between the transgene and the endogenous gene expression, each of the four candidate genes was amplified from Golden Promise and the sequence was compared to the Cebada Capa alleles. Single nucleotide polymorphisms (SNPs) identified between the two allelic sequences were then used to design primers specific for the Cebada Capa transgenes (Table 6-1). Primer pairs were designed in distinct exons so that the resulting RT-PCR product could be distinguished from a fragment amplified from contaminating genomic DNA (Figure 6-1). Cebada Capa and transgenic T_1 plants transformed with *Hvpg1* and *HvHGA1* amplified an 830 bp and a 192 bp fragment, respectively, whereas no product was obtained from non-transgenic plants and Golden Promise (Figure 6-4A). A second pair of primers at the end of the last exon of *Hvpg1* (Figure 6-1) amplified a 623 bp fragment in Cebada Capa and in transgenic T_1 plants, confirming that the complete *Hvpg1* gene is present in the transformed plants. Amplification with these primers resulted in a ~1.4 kb fragment in Golden Promise due to an insertion of ~800 bp corresponding to additional terminal repeats in the Gly-rich repeat domain compared to Cebada Capa (see chapter V). The primers designed for the *Hvpg4* and *HvHGA2* genes were not specific for the Cebada Capa alleles as amplification products were also obtained with Golden Promise DNA. Therefore, a nested PCR strategy was used. A first set of non-specific primers were used to amplify a 390 bp fragment for *Hvpg4* and a 610 bp fragment for *HvHGA2* (Table 6-1 and Figure 6-4B). The PCR products were subsequently used as templates for a nested PCR with Cebada Capa-specific primers (Table 6-1). The specific primers amplified a 278 bp fragment for *Hvpg4* and a 460 bp fragment for *HvHGA2* only from mRNA of transgenic plants and from Cebada Capa (Figure 6-4B). Thus, the RT-PCR experiments demonstrated that all four transgenes are expressed

in the T_1 plants and that the absence of resistance reaction after leaf rust infection is not due to a lack of transgene expression.

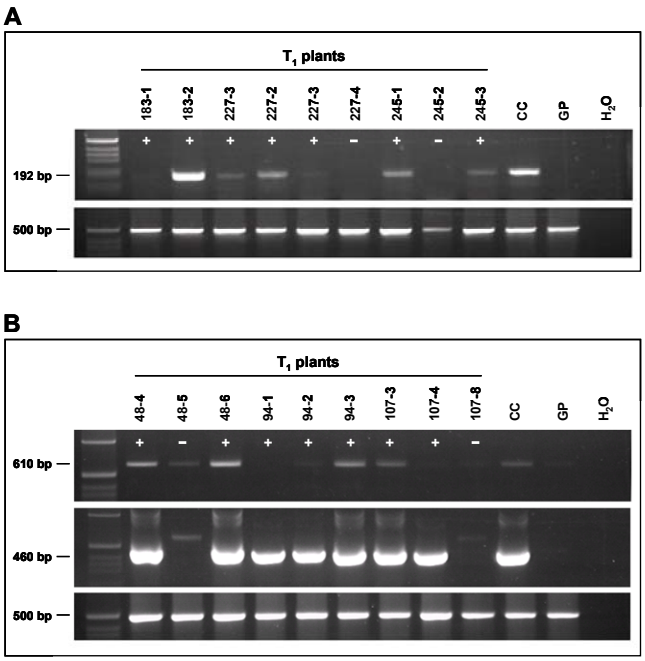


Figure 6-4 RT-PCR analysis demonstrating expression of the transgenes in T_1 plants

RT-PCR was performed on mRNA extracted from transgenic (+) and non-transgenic (-) T_1 plants based on Southern data. Control reactions with mRNA from Cebada Capa and Golden Promise are indicated as CC and GP, respectively. Primers amplifying a 500 bp fragment of the GAP-DH gene were used as a positive control for PCR amplification (lower panels). (A) RT-PCR with primers specific for the Cebada Capa allele of *HvHGA1* (upper panel). (B) RT-PCR with non-specific primers (upper panel) and nested PCR with primers specific for the Cebada Capa allele of *HvHGA2* (middle panel).

3.3 Virus-induced gene silencing of the two candidate genes *Hvpg1* and *HvHGA1*

Virus-induced gene silencing (VIGS) was used to accelerate the validation process of the candidate genes and to assess its potential as a method of identification for barley leaf rust resistance genes. At the time of the experiment, transgenic plants for the *Hvpg4* and *HvHGA2* genes were already analyzed. Therefore, the VIGS experiment was performed with the two remaining candidates *Hvpg1* and *HvHGA1*. We first tested the ability of the *Barley stripe mosaic virus* (BSMV) to move systemically in the cultivars Cebada Capa and Morex by targeting the phytoene desaturase (*PDS*) gene which provides a convenient visual reporter for

silencing (Holzberg *et al.*, 2002). Photobleaching as the result of suppression of *PDS* activity was observed in both cultivars, demonstrating the capacity of BSMV to spread in these lines (Figure 6-5A). Seven-day-old Cebada Capa seedlings were inoculated with a 1:1:1 mixture of *in vitro* transcripts synthesized from plasmids containing the wild-type BSMV α and β RNAs and derivates of the γ RNA that carried one of three cDNA fragments amplified from *Hvpga1* (BSMV:pg1-1, BSMV:pg1-2, and BSMV:pg1-3) or a fragment of *HvHGA1* (BSMV:hga1-1) (Figure 6-5B). Morex seedlings were inoculated with wild-type viral RNAs that carried no plant sequence (BSMV:00). Ten or 13 days after BSMV inoculation, plants were artificially inoculated with the barley leaf rust isolate 1.2.1 which is avirulent on *Rph7*. Infection of Morex with the control BSMV:00 transcript had no effect on the susceptible phenotype displayed by this cultivar, indicating that virus infection does not interfere with the development of the fungal pathogen. The resistance reaction of Cebada Capa upon leaf rust infection was not altered in any of the plants inoculated with the different BSMV constructs for *Hvpg1* and *HvHGA1*. These data indicate that neither *Hvpg1* nor *HvHGA1* correspond to the *Rph7* resistance gene and confirm the results obtained from the stable transformation experiments (see 3.1), *i.e.* none of the transgenic plants expressing *Hvpg1* and *HvHGA1* showed resistance to leaf rust *AvrRph7*.

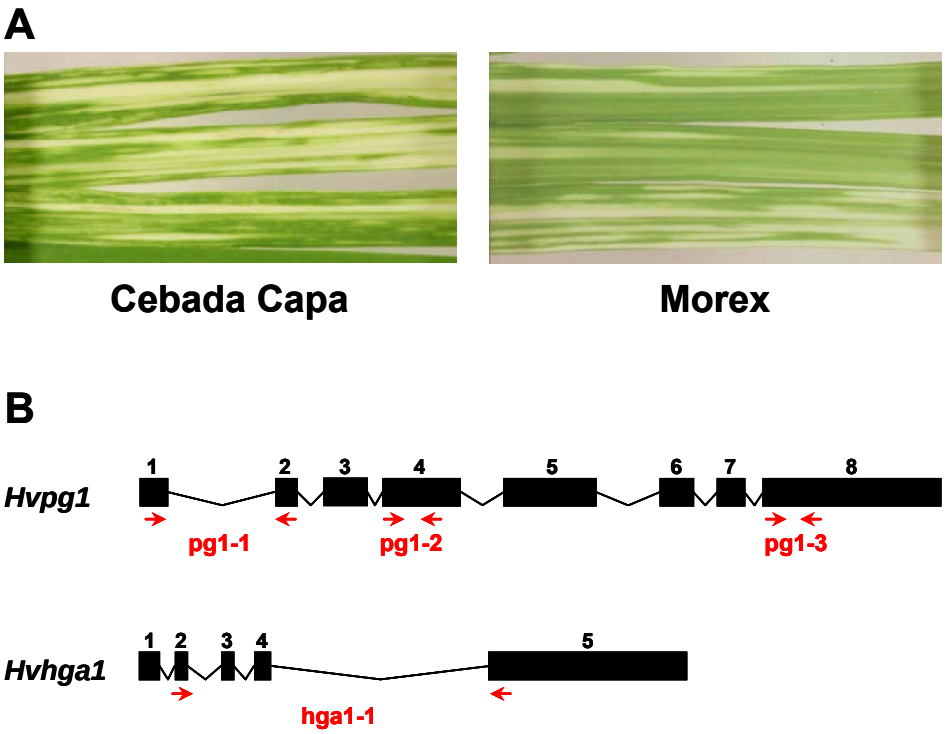


Figure 6-5 VIGS experiment with the two candidate genes *Hvpg1* and *Hvhga1*

(A) Silencing of *PDS* in barely cv. Cebada Capa and Morex by BSMV. Plants were infected with *in vitro* transcribed RNAs from a BSMV construct in which the γ RNA was engineered to carry a 185 bp fragment of *PDS*. Supression of PDS resulted in photobleaching, demonstrating the ability of BSMV to spread in both cultivars. (B) Schematic representation of the two genes *Hvpg1* and *Hvhga1*. Primers used to amplify cDNA fragments for the different VIGS constructs are indicated with red arrows.

3.4 Identification of γ-irradiation *Rph7* mutants

In addition to the stable transformation and the VIGS experiment, a forward genetic approach was undertaken to identify *Rph7* among the candidate genes. 70,000 plants of a γ-ray mutagenized M_2 population originating from a double haploid (DH) Bowman/8* Cebada Capa near-isogenic line (NIL) carrying *Rph7* in the background of the susceptible line Bowman, were phenotypically screened after artificial leaf rust infection. Seven susceptible plants were identified. For four of them,

the susceptible phenotype was confirmed in the M_3 generation. Two showed a phenotype similar to the one observed in Bowman whereas the other two showed brown-grey necroses with very small pustules in the middle which is an intermediate phenotype between Cebada Capa and Bowman. Southern hybridization with the *Rph7* flanking markers XHv480 and XHvgad1 (see 3.5) and with specific RFLP probes for the *HvHGA1, HvHGA2,* and *Hvpg1* genes showed that the γ-ray mutagenesis did not cause large deletions and that all mutants originated from the resistant NIL background and did not result from out-crossing or contamination events (data not shown). These four plants were then back-crossed with Bowman to determine whether the mutation is in the *Rph7* gene or in a gene that acts downstream in the resistance signalling pathway. If all F_1 progeny plants are susceptible, the mutation is expected to be in the *Rph7* gene. However, if the F_1 progeny segregate, another gene is mutated. In none of the four identified mutants the *Rph7* gene was affected, as resistant or intermediate resistant plants were detected in the progeny of all four back-crosses. Two mutant F_2 plants which carry homozygous alleles for Cebada Capa at the *Rph7* locus (based on Southern hybridization with a probe corresponding to the *Hvpg1* gene) were selected for further cross with Cebada Capa to determine the genetic localization of the gene affected by the mutation.

Thus, no *Rph7* mutant was identified. A possible explanation for this result is that mutants with a larger deletion at the *Rph7* locus could be lethal as the deletion might comprise a vital gene. Large deletions on chromosomes were reported for γ-ray mutagenesis in plants, compared to ethyl methanesulfonate (EMS) treatment (Koebner and Hadfield, 2001). However, small deletion events have also been found for γ-ray *Lr10* and *Pm3b* mutants (Feuillet *et al.*, 2003; Yahiaoui *et al.*, 2004). The

attempts to produce an EMS mutagenized *Rph7* mutant M_2 population failed so far. However, we have generated an alternative M_2 population using fast neutron irradiation that usually triggers smaller deletions than γ-ray irradiation (100–200 kb). This population could be screened in the future.

3.5 Possible reasons for the lack of complementation with the four candidate genes and perspectives

The complementation experiments with the four candidate genes *Hvpg1, Hvpg4, HvHGA1,* and *HvHGA2* as well as the VIGS assay did not lead to the identification of *Rph7*. Different hypotheses can be considered to explain these results.

Genetic mapping of *Rph7* was performed in a population consisting of 1,157 F_2 plants derived from a cross between the resistant cv. Cabada Capa and the susceptible cv. Bowman (Brunner *et al.*, 2003). Out of the 1,157 F_2 plants, a subpopulation of 105 recombinant F_2 plants was selected for high-resolution mapping and the phenotype of each of these 105 F_2 plant was confirmed by artificial leaf rust infection of 15 F_3 seedlings. *Rph7* was mapped within an interval of 0.13 cM representing ~200 kb in Cebada Capa between the flanking markers XHvgad1, XHv283, and XHv480 (Figure 6-6). The low-pass shotgun marker Hvgad1 mapped at two recombination events distal to *Rph7* (F_2 plants #389 and #935) whereas the two markers Hv283 and Hv480 were found at one recombination event proximal to it (F_2 plant #1150) (Brunner *et al.*, 2003). During the project, we have obtained homologous sequences from Cabada Capa and Morex spanning the *Rph7* locus (chapter V). This allowed us to develop new genomic sequence tagged sites (STS)

markers in order to re-confirm the recombination events and the genetic location of *Rph7*. Based on the sequence polymorphisms between Cebada Capa and Morex, primers were designed that amplify polymorphic fragments near the recombination breakpoints that were originally analyzed using the flanking RFLP markers (Figure 6-6).

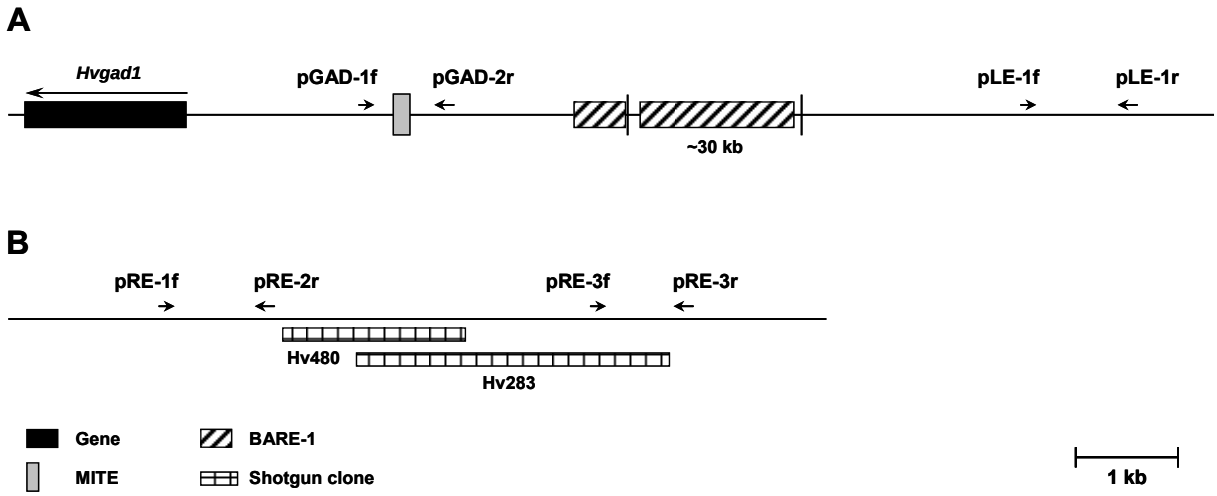


Figure 6-6 Confirmation of the recombination breakpoints at the *Rph7* locus in Cebada Capa

Genomic region around the distal (**A**) and proximal (**B**) recombination breakpoints in Cebada Capa. Flanking RFLP probes are shown as black (Hvgad1) and checkered (Hv480 and Hv283) boxes, respectively. Positions of the primers used to amplify polymorphic STS markers are indicated with arrowheads.

Ten (plant #389), 24 (plant #935), and 35 (plant #1150) F_4 seedlings originating from the three previously identified recombinant F_2 plants were artificially infected with leaf rust and the resistance phenotype was determined for each plant. PCR was performed on genomic DNA with the primers described in Table 6-3. The marker LE-1f/1r cosegregated with *Rph7*. However, the marker GAD-1f/2r identified the

recombination event in the plants #389 and #935, confirming the recombination breakpoint previously identified by XHvgad1 distal to the *Rph7* gene. Similarly, the marker RE-1f/2r was linked to *Rph7* whereas marker RE-3f/3r confirmed the proximal recombination event in plant #1150. These results demonstrate that the previous genetic mapping of *Rph7* was correct and therefore indicates that the *Rph7* gene is present in the interval between the flanking markers. Currently, the physical distance between the recombination breakpoints is ~200 kb. To further narrow down this interval, the mapping population will need to be increased by 2,000 to 3,000 additional F_2 plants. New genomic STS markers can be generated based on the homologous Cabada Capa and Morex BAC sequences. These markers should help identifying further recombination events and thus help to delimitate the location of the *Rph7* gene in a smaller physical interval.

The sequence annotation and the gene predictions that identified the candidate genes for *Rph7* were performed in 2002 (Brunner *et al.*, 2003). Gene annotation is constantly evolving with the release of new Expressed Sequence Tag (EST) sequences and the availability of improved prediction algorithms. It is therefore possible that the automated gene prediction program [RiceGAAS: http://ricegaas.dna.affrc.go.jp/] and the Blast analysis used in 2002 were not able to detect a small gene present in the intergenic regions between the four candidate genes because of its size, structure, and lack of matching EST in the databases. We have now re-annotated the Cebada Capa and Morex BAC sequences using the current EST and Triticeae Repeat Sequence [TREP http://wheat.pw.usda.gov/ITMI/Repeats/index.shtml] databases. This did not lead to the identification of a new or previously not annotated gene. However, there is more and more evidence that small DNA sequences, which cannot be detected by current prediction pipelines and do not

match EST databases, do have a functional role. Moreover, recent literature reports that non-coding RNAs, such as microRNAs (miRNAs) and small interfering RNAs (siRNAs) are involved in developmental regulation, genome maintenance, and defence in eukaryotes (Du and Zamore, 2005). miRNAs are a class of recently discovered non-coding small RNAs, usually consisting of ~20-24 nt for plants. Many plant miRNAs are evolutionarily conserved from species to species, some from angiosperms to mosses (reviewed in Chen, 2005; Zhang *et al.*, 2006). Although transcription factor-encoding miRNA targets are the best characterized so far, miRNA targets encoding proteins with various other cellular functions, such as protein degradation, metabolism and stress responses, have been predicted and miRNAs probably also play a role in disease resistance in plants. For example, recent work has demonstrated the involvement of miRNAs in pathogen-induced, post-transcriptional gene silencing (PTGS) as an innate antiviral defense mechanism in higher plants (Ding, 2000).

VII. GENERAL DISCUSSION

1. Comparative genomics in wheat and barley as a tool to reveal mechanisms of genome evolution in the Triticeae

In the last years, a number of large insert BAC libraries have been developed for important grass species such as maize, rice, sorghum, barley, and wheat [http://www.genome.clemson.edu/groups/bac/]. BAC sequencing projects have yielded a large amount of genomic DNA data, allowing to perform microcolinearity studies, *i.e.* comparing the organization of orthologous loci among different grass genomes at the molecular level. Apart from getting an insight into the genomic structure of orthologous loci, these studies have also revealed some of the mechanisms that have shaped the different grass genomes during the last 50-60 million years of evolution (reviewed *e.g.* in Feuillet and Keller, 2002; Bennetzen and Ma, 2003). Many small-scale genomic rearrangements, such as retroelement and gene insertions, deletions, duplications, or inversions were found to have contributed to grass genome evolution. Microcolinearity studies in the grasses have mainly compared orthologous regions in different species, *e.g.* of maize, sorghum, and rice. Only a few studies have performed comparisons at the intraspecific level, that is, have compared homologous sequences from different lines or varieties of the same species. In maize, sequence comparisons of allelic regions between inbred lines has demonstrated that violation of colinearity between two lines of the same species can be as drastic as between different species (Fu and Dooner, 2002; Song and Messing, 2003; Brunner *et al.*, 2005). Moreover, with the discovery of the role of helitron transposable elements in gene rearrangement (see below), these studies

have also revealed new mechanisms of genome evolution. It was speculated that the complementation of haplotypes carrying different non-shared genes could contribute to the phenomenon of heterosis in maize (Fu and Dooner, 2002). Brunner *et al.* (2005) suggested an alternative hypothesis for the role of non-shared sequences in heterosis, focusing on the differences in the repetitive fraction rather than in the genes. Several of the conserved and active genes in the two inbreds Mo17 and B73 are flanked by different retroelements and the different repetitive sequence environments likely affect tissue specificity or temporal regulation of the expression of neighboring genes. Such differences have been proposed to be the cause for heterotic complementation (Song and Messing, 2003).

In this study, we have performed structural analyses at the molecular level of two leaf rust *R* gene loci in wheat and barley. The *Lr10* locus was analyzed in the homoeologous A genomes of diploid, tetraploid, and hexaploid wheat species and the *Rph7* locus was compared between two barley cultivars. Whereas the A genomes of wheat are estimated to have diverged from a common ancestor about 0.5-3 mya (Huang *et al.*, 2002; Wicker *et al.*, 2003), the two barley landraces have probably separated less than 0.5 mya. Thus, we have studied the evolution of two *R* gene loci (against a similar fungal pathogen) at different genomic levels and at different time scales. These comparisons revealed several mechanisms that played a role during the evolution of Triticeae genomes and provided insight into the haplotype structure of *R* loci in these species.

In both cases, the structure of the *R* locus does not correspond to a cluster of genes as it has been observed at most disease *R* loci in plants (*e.g.* the *N* locus of tobacco, the Arabidopsis *Rpp1 and Rpp5* loci, the *Cf-4/9* locus in tomato, or the barley *Mla* locus). The *Lr10* gene codes for a CC-NBS-LRR type of protein while

Rph7, even if not yet identified, does not correspond to this classical type of *R* gene. Thus, the type of *R* gene for the same disease (leaf rust) is not conserved in the two related species wheat and barley.

A large deletion/inversion event has eliminated the resistance gene from the susceptible H2 haplotype at the *Lr10* locus. In this respect, the *Lr10* locus is similar to the bacterial resistance gene locus *Rpm1* in Arabidopsis (Grant *et al.*, 1998). The *Rpm1* locus also exists as two stable haplotypes in Arabidopsis and susceptibility is caused by the complete absence of *Rpm1* (*rpm1-null*). The presence of *rpm1-null* loci with no significant similarity to the Arabidopsis *rpm1-null* alleles in *Brassica napus* suggests that *Rpm1* evolved before the divergence of the Brassicaceae, and that it has been deleted independently in the *Brassica* and Arabisopsis lineages. At the *Rph7* locus, however, no large differences in gene content were found between the susceptible A1B1 haplotype of Morex and the resistant A2B2 Cebada Capa haplotype. Six genes were conserved in order and orientation between the two haplotypes. Indeed, a single gene previously identified in Morex that was absent from the Cebada Capa sequence is part of a helitron transposable element (see below).

In total, six different haplotype combinations were found at the *Rph7* locus among the 41 barley cultivars that have separated probably less than 0.5 mya. To determine to which extent the observed haplotype diversity originates from the wild barley pool, we have extended our analysis to a collection of 147 wild (*Hordeum spontaneum*) lines from Israel (data not shown). Eight haplotypes were found in the wild pool, of which four are in common with the ones identified in the cultivated lines. Thus, two haplotypes that are not present in the wild pool were observed in the cultivated pool while four are unique to the wild pool.

Sequence comparison between H1 haplotypes derived from two different wheat species (*T. monococcum* and *T. turgidum* ssp *durum*) revealed a very different composition of intergenic regions which mainly consist of blocks of repetitive elements. Still, several elements were found at conserved positions on the two sequences despite 2.9-3.3 million years of divergence between the two species (Wicker *et al.*, 2003). In contrast, all the repetitive elements present in the intervals separating the seven conserved regions at the *Rph7* locus were completely different between the two barley sequences.

Together, this study has revealed different types of haplotype conservations at the *Lr10* and *Rph7* locus. Whereas the two *Lr10* haplotypes are of ancient origin and were very stable during evolution and throughout polyploidization events in the last 0.5-3 million years of A genome divergence, the high haplotype diversity at the *Rph7* locus is indicative of a very dynamic barley genome in recent evolutionary times.

Thus, our analyses demonstrated that intraspecific sequence comparison can be as useful as comparisons between different species to identify mechanisms of genome evolution. Based on our findings in barley and the observations made in maize it can be expected that similar disruptions of microcolinearity might also be found between different wheat varieties. However, no comparative analysis has been performed yet between different wheat varieties of the same ploidy level. BAC libraries from the hexaploid *T. aestivum* cultivars Chinese Spring, Renan and Glenlea are now available. This opens the possibility to perform similar types of studies in wheat. It will be interesting to compare the rate and mechanisms of evolution at a similar time scale in barley and wheat, two species that have very different population histories. Hexaploid wheat resulted from two polyploidization events and genetic diversity in wheat was strongly reduced by breeding selection. Barley, however, is

not polyploid and genetic diversity was kept high due to introgressions from wild barley (*Hordeum spontaneum*) during breeding.

2. A helitron transposable element is present at the *Rph7* locus in Morex

In recent years, intraspecific comparative sequencing uncovered dramatic non-colinearity among maize inbred lines (Fu and Dooner, 2002; Song and Messing, 2003; Brunner *et al.*, 2005). Although the majority of the differences were due to recent LTR-retrotransposon insertions, surprisingly enough, the lack of colinearity was not restricted to repetitive elements but also concerned genes or gene fragments (pseudogenes). Homologous BAC sequence comparison between five genomic regions of the Mo17 and B73 inbred lines have shown that many of the pseudogenes are clustered and tend to have the same orientation with respect to the direction of transcription (Brunner *et al.*, 2005). In total, 23 (34%) of the 68 identified genes were not shared between Mo17 and B73. The genic sequences present in one but not the other inbred line (non-shared gene) seem to result from insertion events rather than from deletions as they were, unlike most of the shared genes, consistently not found in the orthologous rice genomic region. The high frequency of non-shared genic fragments raised the question of the molecular mechanism of these insertions. Recently, by comparing all genetic insertion termini among themselves and with their insertion sites, Morgante *et al.* (2005) have found that pseudogene clusters in maize are parts of non-autonomous helitron elements. Helitrons are a new class of eukaryotic transposable elements that were found by computational analysis of genomic sequences from Arabidopsis, rice, and *Caenorhabditis elegans* (Kapitonov

and Jurka, 2001). The putative autonomous elements are large (5.5-15 kb) and encode a signature protein (HEL) composed of the rolling-circle replication initiator and DNA helicase domains necessary for transposition. Plant elements also contain a replication protein A (RPA)-like protein with putative ssDNA-binding activity (Kapitonov and Jurka, 2001). Unlike other DNA transposons, helitrons do not have terminal inverted repeats and do not cause target site duplications. Instead, they have 5'-TC and 3'-CTRR termini preceded by an 18-25 bp region capable of forming a hairpin structure and insert into an A↓T target site. Rolling-circle replication seems a likely model for helitron transposition (Kapitonov and Jurka, 2001). It is proposed that helitrons copy and duplicate pieces of genes and assemble them together for transposition to a new genomic location, while the ancestral copy is left in place. Several non-autonomous maize helitron elements carrying multiple gene fragments were analyzed by Morgante *et al.* (2005) and Brunner *et al.* (2005) in detail. RT-PCR experiments demonstrated that some of the non-autonomous helitrons produce chimeric transcripts bringing exons derived from different genes together within the helitron. This led to the speculation that, during evolution, transcribed helitrons containing several gene fragments may occasionally give rise to new genes with novel functions by a combinatorial assembly of exons. Thus, helitrons may also be involved in the evolution of gene function.

At the *Rph7* locus, we had identified a large insertion of 22 kb containing a putative helicase gene (*Hvhel1*) in Morex that was not present in the allelic Cebada Capa sequence (see chapter V). Re-analysis of the sequence revealed that this helicase gene is part of a helitron transposon. The 5' end of the 22 kb insertion is intact and shows the characteristic TC-terminus. At the 3' end, however, a part of the helitron and additional 300 bp that are present in Cebada Capa adjacent to the

insertion position are deleted in Morex. No additional genes or gene fragments could be identified in the remaining inserted sequence. Thus, this helitron transposon does not seem to have acquired and transposed fragments of genes originating from another region of the barley genome. The presence of the helicase gene, however, suggests that it is an autonomous helitron element. Whether the helitron insertion at the *Rph7* locus in Morex has any implication on the resistance/susceptibility to leaf rust is not clear at the moment. The presence/absence of the 22 kb insertion, together with two other small insertions/duplications, defined four different haplotypes (B1-B4) in the CR4-CR5 interval. The insertion of the helitron is conserved only in lines with the A1B1 haplotype that is, in nine out of the 41 barley cultivars. These nine lines include Morex, Sundance, Franka, Bowman, and five near-isogenic lines carrying different *Rph* genes in the background of Bowman. Additional comparisons of the *Rph7* region in other barley cultivars and further comparative studies at other loci are needed to define whether helitrons have such a predominant role in barley as they have in maize. In wheat, no traces of helitron elements have been identified on the available sequences so far. However, sequence comparisons between wheat lines of different ploidy levels are probably not the right tool to identify helitrons as they might have diverged too much to be recognized anymore. Microcolinearity studies between different wheat varieties of the same ploidy level might shed light on this question.

3. Future prospects

As discussed above (chapter IV, 3.5), to further progress in the isolation of *Rph7* it will be necessary to increase the mapping population by 2,000 to 3,000 additional F_2 plants. The current physical distance between the recombination breakpoints in Cebada Capa is ~200 kb. Transformation with complete BAC clones has not yet been demonstrated to be efficient in barley. Therefore, the interval containing *Rph7* has to be narrowed down to perform complementation experiments with the complete genomic region between flanking markers. The two newly developed genomic STS markers can be used to identify additional recombinant plants in the current interval. To localize the recombination breakpoints within the interval, it will be necessary to generate new markers based on the homologous Cebada Capa and Morex BAC sequences. It is assumed that the high level of sequence divergence at the *Rph7* locus affects recombination in this region. Only conserved sequences can engage in homologous recombination in the cross between Cebada Capa and Bowman. Thus, primers for new STS markers should be designed in conserved regions, *i.e.* within genes or in the conserved regions surrounding the genes.

It should be possible to perform *Agrobacterium*-mediated transformation with fragments of about 20 kb. Once a 20 kb sequence that is able to complement the susceptible Golden Promise phenotype will have been identified, further complementation experiments with smaller sub-fragments will be necessary. The 20 kb fragment could be sub-cloned after digestion or shearing. However, *Agrobacterium*-mediated transformation is labor- and time-intensive. Alternatively,

VIGS experiments with 300 bp fragments covering the 20 kb sequence could be used to identify the stretch of DNA harboring *Rph7*.

Once *Rph7* will have been isolated, it will be interesting to identify genes functioning in leaf rust resistance mediated by this new type of *R* gene. Similarly to the strategy applied by Scofield *et al.* (2005) to analyze the *Lr21*-mediated resistance pathway, VIGS could be used to assay genes (such as *Rar1, Sgt1*, and *Hsp90*) known to be required in many resistance pathways. Additionally, the two susceptible mutants identified to be affected in a gene different from *Rph7* might be valuable to reveal genes in the resistance signaling pathway.

BAC libraries from additional grass species such as Brachypodium, wheat, and rye grass are now available, allowing to perform further evolutionary studies at the *Rph7* locus. Data from these species are expected to provide information about the origin of the gene movements that, as we suggest, have led to the different situations observed in rice and the Triticeae lineage.

VIII. REFERENCES

Adams, K.L., and Wendel, J.F. (2005). Novel patterns of gene expression in polyploid plants. Trends Genet 21, 539-543.

Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. Curr Opin Plant Biol 8, 135-141.

Allouis, S., Moore, G., Bellec, A., Sharp, R., Faivre Rampant, P., Mortimer, K., Pateyron, S., Foote, T.N., Griffiths, S., Caboche, M., and Chalhoub, B. (2003). Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L.) BAC library from the reference germplasm 'Chinese Spring'. Cereal Res Commun 31, 331-338.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Ananiev, E.V., Phillips, R.L., and Rines, H.W. (1998). Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. Genetics 149, 2025-2037.

Badr, A., Muller, K., Schafer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H.H., Pozzi, C., Rohde, W., and Salamini, F. (2000). On the origin and domestication history of Barley (*Hordeum vulgare*). Mol Biol Evol 17, 499-510.

Baumgarten, A., Cannon, S., Spangler, R., and May, G. (2003). Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics 165, 309-319.

Bennett, M.D., and Leitch, I.J. (1995). Nuclear-DNA amounts in angiosperms. Ann Bot 76, 113-176.

Bennetzen, J.L. (2000). Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. Plant Cell 12, 1021-1029.

Bennetzen, J.L. (2002). Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115, 29-36.

Bennetzen, J.L. (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15, 621-627.

Bennetzen, J.L., and Ramakrishna, W. (2002). Exceptional haplotype variation in maize. Proc Natl Acad Sci USA 99, 9093-9095.

Bennetzen, J.L., and Ramakrishna, W. (2002). Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 48, 821-827.

Bennetzen, J.L., and Ma, J. (2003). The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6, 128-133.

Bennetzen, J.L., Ma, J., and Devos, K.M. (2005). Mechanisms of recent genome size variation in flowering plants. Ann Bot 95, 127-132.

Bergelson, J., Kreitman, M., Stahl, E.A., and Tian, D. (2001). Evolutionary dynamics of plant *R*-genes. Science 292, 2281-2285.

Bieri, S., Mauch, S., Shen, Q.H., Peart, J., Devoto, A., Casais, C., Ceron, F., Schulze, S., Steinbiss, H.H., Shirasu, K., and Schulze-Lefert, P. (2004). RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. Plant Cell 16, 3480-3495.

Blanc, G., and Wolfe, K.H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16, 1679-1691.

Blanc, G., Hokamp, K., and Wolfe, K.H. (2003). A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res 13, 137-144.

Bonfield, J.K., Smith, K., and Staden, R. (1995). A new DNA sequence assembly program. Nucleic Acids Res 23, 4992-4999.

Brueggeman, R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., Druka, A., Steffenson, B., and Kleinhofs, A. (2002). The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99, 9328-9333.

Brunner, S., Keller, B., and Feuillet, C. (2000). Molecular mapping of the *Rph7.g* leaf rust resistance gene in barley (*Hordeum vulgare* L.). Theor Appl Genet 101, 783-788.

Brunner, S., Keller, B., and Feuillet, C. (2003). A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. Genetics 164, 673-683.

Brunner, S., Pea, G., and Rafalski, A. (2005). Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize. Plant J 43, 799-810.

Brunner, S., Fengler, K., Morgante, M., Tingey, S., and Rafalski, A. (2005). Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17, 343-360.

Buschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., vanDaelen, R., vanderLee, T., Diergaarde, P., Groenendijk, J., Topsch, S., Vos, P., Salamini, F., and Schulze-Lefert, P. (1997). The barley *mlo* gene: A novel control element of plant pathogen resistance. Cell 88, 695-705.

Caldwell, K.S., Langridge, P., and Powell, W. (2004). Comparative sequence analysis of the region harboring the hardness locus in barley and its colinear region in rice. Plant Physiol 136, 3177-3190.

Cenci, A., Chantret, N., Kong, X., Gu, Y., Anderson, O.D., Fahima, T., Distelfeld, A., and Dubcovsky, J. (2003). Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor Appl Genet 107, 931-939.

Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdille, P., Joudrier, P., Gautier, M.F., Cattolico, L., Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P., and Chalhoub, B. (2005). Molecular basis of evolutionary events that shaped the *Hardness* locus in diploid and polyploid wheat species (Triticum and Aegilops). Plant Cell 17, 1033-1045.

Chao, S., Sharp, P.J., and Gale, M.D. (1988). A linkage map of wheat homoeologous group of chromosomes using RFLP markers. In Proceedings of the 7th International Wheat Genetic Symposium IPSR, T.E. Miller and R.M.D. Koebner, eds (Cambridge, UK: Institute of Plant Science Research, Cambridge Laboratory), pp. 493-498.

Charlesworth, D., Bartolome, C., Schierup, M.H., and Mable, B.K. (2003). Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. Mol Biol Evol 20, 1741-1753.

Chen, M., SanMiguel, P., and Bennetzen, J.L. (1998). Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 148, 435-443.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.S., Zhang, H., Wing, R.A., and Bennetzen, J.L. (1997). Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. Proc Natl Acad Sci USA 94, 3431-3435.

Chen, X.M. (2005). microRNA biogenesis and function in plants. Febs Lett 579, 5923-5931.

Collins, N.C., Lahaye, T., Peterhansel, C., Freialdenhoven, A., Corbitt, M., and Schulze-Lefert, P. (2001). Sequence haplotypes revealed by sequence-tagged site fine mapping of the *Ror1* gene in the centromeric region of barley chromosome 1H. Plant Physiol 125, 1236-1247.

Dangl, J.L., and Jones, J.D. (2001). Plant pathogens and integrated defence responses to infection. Nature 411, 826-833.

De Bustos, A., Loarce, Y., and Jouve, N. (2002). Species relationships between antifungal chitinase and nuclear rDNA (internal transcribed spacer) sequences in the genus *Hordeum*. Genome 45, 339-347.

Devos, K.M. (2005). Updating the 'Crop Circle'. Curr Opin Plant Biol 8, 155-162.

Devos, K.M., and Gale, M.D. (2000). Genome relationships: The grass model in current research. Plant Cell 12, 637-646.

Devos, K.M., Brown, J.K., and Bennetzen, J.L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12, 1075-1079.

DeWet, J.M.J. (1986). Hybridization and polyploidy in the Poaceae. In Grass systematics and evolution, T. Soderstrom, K.W. Hilu, C.S. Campell, and M.E. Barkworth, eds (Washington DC, USA: Smithsonian Institution Press), pp. 188-194.

Ding, S.W. (2000). RNA silencing. Curr Opin Biotechnol 11, 152-156.

Du, T.T., and Zamore, P.D. (2005). microPrimer: The biogenesis and function of microRNA. Development 132, 4645-4652.

Dubcovsky, J., Ramakrishna, W., SanMiguel, P.J., Busso, C.S., Yan, L., Shiloff, B.A., and Bennetzen, J.L. (2001). Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 125, 1342-1353.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175-185.

Feldman, M. (2001). Origin of cultivated wheat. In The world wheat book: A history of wheat breeding, A.P. Bonjean and W.J. Angus, eds (Paris, France: Lavoisier.

Feuillet, C., and Keller, B. (1999). High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96, 8265-8270.

Feuillet, C., and Keller, B. (2002). Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. Ann Bot 89, 3-10.

Feuillet, C., Travella, S., Stein, N., Albar, L., Nublat, A., and Keller, B. (2003). Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100, 15253-15258.

Flor, H. (1971). Current status of the gene-for-gene concept. Annu Rev Phytopathol 9, 275-296.

Foote, T.N., Griffiths, S., Allouis, S., and Moore, G. (2004). Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum*: Its use as a tool to bridge the gap between rice and wheat in elucidating gene content. Funct Integr Genomics 4, 26-33.

Fu, H., and Dooner, H.K. (2002). Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99, 9573-9578.

Gale, M.D., and Devos, K.M. (1998). Comparative genetics in the grasses. Proc Natl Acad Sci USA 95, 1971-1974.

Gaut, B.S. (2002). Evolutionary dynamics of grass genomes. New Phytol 154, 15-28.

Gaut, B.S., and Clegg, M.T. (1993). Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc Natl Acad Sci USA 90, 5095-5099.

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93, 10274-10279.

Gaut, B.S., d'Ennequin, M.L., Peek, A.S., and Sawkins, M.C. (2000). Maize as a model for the evolution of plant nuclear genomes. Proceedings of the National Academy of Sciences of the United States of America 97, 7008-7015.

Graner, A., Bjornstad, A., Konishi, T., and Ordon, F. (2003). Molecular diversity of the barley genome. In Diversity in Barley (*Hordeum vulgare*), R.V. Bothmer, T.V. Hintum, H. Knüpffer, and K. Sato, eds (Amsterdam: Elsevier Science), pp. 122-141.

Graner, A., Siedler, H., Jahoor, A., Herrmann, R.G., and Wenzel, G. (1990). Assessment of the degree and the type of restriction fragment length polymorphism in barley (*Hordeum vulgare*). Theor Appl Genet 80, 826-832.

Grant, M.R., McDowell, J.M., Sharpe, A.G., de Torres Zabala, M., Lydiate, D.J., and Dangl, J.L. (1998). Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. Proc Natl Acad Sci USA 95, 15843-15848.

Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L. (1995). Structure of the *Arabidopsis Rpm1* gene enabling dual specificity disease resistance. Science 269, 843-846.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.L., White, F.F., and Yin, Z. (2005). *R* gene expression induced by a type-III effector triggers disease resistance in rice. Nature 435, 1122-1125.

Gu, Y.Q., Coleman-Derr, D., Kong, X., and Anderson, O.D. (2004). Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. Plant Physiol 135, 459-470.

Gu, Y.Q., Anderson, O.D., Londeore, C.F., Kong, X., Chibbar, R.N., and Lazo, G.R. (2003). Structural organization of the barley D-hordein locus in comparison with its orthologous regions of wheat genomes. Genome 46, 1084-1097.

Guyot, R., and Keller, B. (2004). Ancestral genome duplication in rice. Genome 47, 610-614.

Guyot, R., Yahiaoui, N., Feuillet, C., and Keller, B. (2004). *In silico* comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct Integr Genomics 4, 47-58.

Halterman, D., Zhou, F., Wei, F., Wise, R.P., and Schulze-Lefert, P. (2001). The MLA6 coiled-coil, NBS-LRR protein confers *AvrMla6*-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. Plant J 25, 335-348.

Halterman, D.A., and Wise, R.P. (2004). A single-amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signaling. Plant J 38, 215-226.

Han, B., and Xue, Y. (2003). Genome-wide intraspecific DNA-sequence variations in rice. Curr Opin Plant Biol 6, 134-138.

Hanson, R.E., Islam-Faridi, M.N., Percival, E.A., Crane, C.F., Ji, Y., McKnight, T.D., Stelly, D.M., and Price, H.J. (1996). Distribution of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105, 55-61.

Holzberg, S., Brosio, P., Gross, C., and Pogue, G.P. (2002). Barley stripe mosaic virus-induced gene silencing in a monocot plant. Plant J 30, 315-327.

Huang, S., Sirikhachornkit, A., Faris, J.D., Su, X., Gill, B.S., Haselkorn, R., and Gornicki, P. (2002). Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant Mol Biol 48, 805-820.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., and Gornicki, P. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99, 8133-8138.

Hulbert, S.H., Webb, C.A., Smith, S.M., and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39, 285-312.

Ilic, K., SanMiguel, P.J., and Bennetzen, J.L. (2003). A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. Proc Natl Acad Sci USA 100, 12265-12270.

Isidore, E., Scherrer, B., Chalhoub, B., Feuillet, C., and Keller, B. (2005). Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res 15, 526-536.

Isidore, E., Scherrer, B., Bellec, A., Budin, K., Faivre-Rampant, P., Waugh, R., Keller, B., Caboche, M., Feuillet, C., and Chalhoub, B. (2005). Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region. Funct Integr Genomics 5, 97-103.

Islam-Faridi, M.N., Childs, K.L., Klein, P.E., Hodnett, G., Menz, M.A., Klein, R.R., Rooney, W.L., Mullet, J.E., Stelly, D.M., and Price, H.J. (2002). A molecular cytogenetic map of sorghum chromosome 1: Fluorescence *in situ* hybridization analysis with mapped bacterial artificial chromosomes. Genetics 161, 345-353.

Johal, G.S., and Briggs, S.P. (1992). Reductase activity encoded by the *HM1* disease resistance gene in maize. Science 258, 985-987.

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A.H. (2000). Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97, 6603-6607.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci USA 98, 8714-8719.

Kawchuk, L.M., Hachey, J., Lynch, D.R., Kulcsar, F., van Rooijen, G., Waterer, D.R., Robertson, A., Kokko, E., Byers, R., Howard, R.J., Fischer, R., and Prufer, D. (2001). Tomato *Ve* disease resistance genes encode cell surface-like receptors. Proc Natl Acad Sci USA 98, 6511-6515.

Keller, B., and Feuillet, C. (2000). Colinearity and gene density in grass genomes. Trends Plant Sci 5, 246-251.

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.J., Ohtsuki, K., Shishiki, T., Otomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., Hayashizaki, Y., and Yasunishi, A. (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 301, 376-379.

Kim, J.S., Klein, P.E., Klein, R.R., Price, H.J., Mullet, J.E., and Stelly, D.M. (2005). Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169, 1169-1173.

Kim, J.S., Childs, K.L., Islam-Faridi, M.N., Menz, M.A., Klein, R.R., Klein, P.E., Price, H.J., Mullet, J.E., and Stelly, D.M. (2002). Integrated karyotyping of sorghum by *in situ* hybridization of landed BACs. Genome 45, 402-412.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16, 111-120.

Klein, P.E., Klein, R.R., Vrebalov, J., and Mullet, J.E. (2003). Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J 34, 605-621.

Klein, P.E., Klein, R.R., Cartinhour, S.W., Ulanch, P.E., Dong, J., Obert, J.A., Morishige, D.T., Schlueter, S.D., Childs, K.L., Ale, M., and Mullet, J.E. (2000). A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Res 10, 789-807.

Koebner, R., and Hadfield, J. (2001). Large-scale mutagenesis directed at specific chromosomes in wheat. Genome 44, 45-49.

Kong, X.Y., Gu, Y.Q., You, F.M., Dubcovsky, J., and Anderson, O.D. (2004). Dynamics of the evolution of orthologous and paralogous portions of a complex locus region in two genomes of allopolyploid wheat. Plant Mol Biol 54, 55-69.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17, 1244-1245.

Lacomme, C., Hrubikova, K., and Hein, I. (2003). Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. Plant J 34, 543-553.

Langham, R.J., Walsh, J., Dunn, M., Ko, C., Goff, S.A., and Freeling, M. (2004). Genomic duplication, fractionation and the origin of regulatory novelty. Genetics 166, 935-945.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet 20, 116-122.

Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A., and Schulze-Lefert, P. (1998). Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 95, 370-375.

Levy, M., Edelbaum, O., and Sela, I. (2004). Tobacco mosaic virus regulates the expression of its own resistance gene *N*. Plant Physiol 135, 2392-2397.

Li, W., and Gill, B.S. (2002). The colinearity of the *Sh2/A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the *Triticeae*. Genetics 160, 1153-1162.

Lijavetzky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R., and Dubcovsky, J. (1999). Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42, 1176-1182.

Liu, B., Vega, J.M., and Feldman, M. (1998). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. Genome 41, 535-542.

Ma, J., and Bennetzen, J.L. (2004). Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101, 12404-12410.

Ma, J., Devos, K.M., and Bennetzen, J.L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14, 860-869.

Martin, G.B., Bogdanove, A.J., and Sessa, G. (2003). Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54, 23-61.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. (1993). Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262, 1432-1436.

McDowell, J.M., and Woffenden, B.J. (2003). Plant disease resistance genes: recent insights and potential applications. Trends Biotechnol 21, 178-183.

Menz, M.A., Klein, R.R., Mullet, J.E., Obert, J.A., Unruh, N.C., and Klein, P.E. (2002). A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol 48, 483-499.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15, 809-834.

Moore, G., Devos, K.M., Wang, Z., and Gale, M.D. (1995). Cereal genome evolution. Grasses, line up and form a circle. Curr Biol 5, 737-739.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., and Rafalski, A. (2005). Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37, 997-1002.

Moullet, O., Zhang, H.B., and Lagudah, E.S. (1999). Construction and characterisation of a large DNA insert library from the D genome of wheat. Theor Appl Genet 99, 305-313.

Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J.N., Noyes, T., Oefner, P.J., Stahl, E.A., and Weigel, D. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30, 190-193.

Ohno, S. (1970). Evolution by Gene Duplication. (London: Springer-Verlag).

Ozkan, H., Levy, A.A., and Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13, 1735-1747.

Pan, Q., Wendel, J., and Fluhr, R. (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50, 203-213.

Patel, M., Johnson, J.S., Brettell, R.I.S., Jacobsen, J., and Xue, G.-P. (2000). Transgenic barley expressing a fungal xylanase gene in the endosperm of the developing grains. Mol Breed 6, 113-124.

Paterson, A.H., Bowers, J.E., and Chapman, B.A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101, 9903-9908.

Paterson, A.H., Bowers, J.E., Peterson, D.G., Estill, J.C., and Chapman, B.A. (2003). Structure and evolution of cereal genomes. Curr Opin Genet Dev 13, 644-650.

Radwan, O., Mouzeyar, S., Nicolas, P., and Bouzidi, M.F. (2005). Induction of a sunflower CC-NBS-LRR resistance gene analogue during incompatible interaction with *Plasmopara halstedii*. J Exp Bot 56, 567-575.

Rafalski, A., and Morgante, M. (2004). Corn and humans: Recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20, 103-111.

Ramakrishna, W., Dubcovsky, J., Park, Y.J., Busso, C., Emberton, J., SanMiguel, P., and Bennetzen, J.L. (2002). Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 162, 1389-1400.

Ramakrishna, W., Emberton, J., SanMiguel, P., Ogden, M., Llaca, V., Messing, J., and Bennetzen, J.L. (2002). Comparative sequence analysis of the sorghum *Rph* region and the maize *Rp1* resistance gene complex. Plant Physiol 130, 1728-1738.

Richly, E., Kurth, J., and Leister, D. (2002). Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol Biol Evol 19, 76-84.

Riely, B.K., and Martin, G.B. (2001). Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. Proc Natl Acad Sci USA 98, 2059-2064.

Rostoks, N., Park, Y.J., Ramakrishna, W., Ma, J., Druka, A., Shiloff, B.A., SanMiguel, P.J., Jiang, Z., Brueggeman, R., Sandhu, D., Gill, K., Bennetzen, J.L., and Kleinhofs, A. (2002). Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct Integr Genomics 2, 51-59.

Sakata, K., Nagamura, Y., Numa, H., Antonio, B.A., Nagasaki, H., Idonuma, A., Watanabe, W., Shimizu, Y., Horiuchi, I., Matsumoto, T., Sasaki, T., and Higo, K. (2002). RiceGAAS: An automated annotation system and database for rice genome sequence. Nucleic Acids Res 30, 98-102.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: A laboratory manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. Nat Genet 20, 43-45.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science 274, 765-768.

Scherrer, B., Keller, B., and Feuillet, C. (2002). Two haplotypes of resistance gene analogs have been conserved during evolution at the leaf rust resistance locus *Lr10* in wild and cultivated wheat. Funct Integr Genomics 2, 40-50.

Schweizer, P., Pokorny, J., Abderhalden, O., and Dudler, R. (1999). A transient assay system for the functional assessment of defense-related genes in wheat. Mol Plant Microbe In 12, 647-654.

Scofield, S.R., Huang, L., Brandt, A.S., and Gill, B.S. (2005). Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the *Lr21*-mediated leaf rust resistance pathway. Plant Physiol 138, 2165-2173.

Shantz, H.L. (1954). The place of grasslands in the earth's cover of vegitation. Ecology 35, 143-145.

Shirasu, K., Schulman, A.H., Lahaye, T., and Schulze-Lefert, P. (2000). A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10, 908-915.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M., and Van de Peer, Y. (2002). The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99, 13627-13632.

Soderlund, C., Humphray, S., Dunham, A., and French, L. (2000). Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 10, 1772-1787.

Song, R., and Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci USA 100, 9055-9060.

Song, R., Llaca, V., and Messing, J. (2002). Mosaic organization of orthologous sequences in grass genomes. Genome Res 12, 1549-1555.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C., and Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270, 1804-1806.

Sonnhammer, E.L., and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167, GC1-10.

Stahl, E.A., Dwyer, G., Mauricio, R., Kreitman, M., and Bergelson, J. (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400, 667-671.

Stein, N., Feuillet, C., Wicker, T., Schlagenhauf, E., and Keller, B. (2000). Subgenome chromosome walking in wheat: A 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). Proc Natl Acad Sci USA 97, 13436-13441.

Stein, N., Perovic, D., Kumlehn, J., Pellio, B., Stracke, S., Streng, S., Ordon, F., and Graner, A. (2005). The eukaryotic translation initiation factor 4E confers multiallelic recessive *Bymovirus* resistance in *Hordeum vulgare* (L.). Plant J 42, 912-922.

Sun, X., Cao, Y., Yang, Z., Xu, C., Li, X., Wang, S., and Zhang, Q. (2004). *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. Plant J 37, 517-527.

Swigonova, Z., Bennetzen, J.L., and Messing, J. (2005). Structure and evolution of the *r/b* chromosomal regions in rice, maize and sorghum. Genetics 169, 891-906.

Tarchini, R., Biddle, P., Wineland, R., Tingey, S., and Rafalski, A. (2000). The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12, 381-391.

Tian, D., Araki, H., Stahl, E., Bergelson, J., and Kreitman, M. (2002). Signature of balancing selection in *Arabidopsis*. Proc Natl Acad Sci USA 99, 11525-11530.

Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M., and Bergelson, J. (2003). Fitness costs of *R*-gene-mediated resistance in *Arabidopsis thaliana*. Nature 423, 74-77.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L., and Avramova, Z. (1999). Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci USA 96, 7409-7414.

Tingay, S., McElroy, D., Kalla, R., Fieg, S., Wang, M.B., Thornton, S., and Brettell, R. (1997). *Agrobacterium tumefaciens*-mediated barley transformation. Plant J 11, 1369-1376.

Tor, M., Brown, D., Cooper, A., Woods-Tor, A., Sjolander, K., Jones, J.D., and Holub, E.B. (2004). *Arabidopsis* downy mildew resistance gene *RPP27* encodes a receptor-like protein similar to CLAVATA2 and tomato Cf-9. Plant Physiol 135, 1100-1112.

Travella, S., Ross, S.M., Harden, J., Everett, C., Snape, J.W., and Harwood, W.A. (2005). A comparison of transgenic barley lines produced by particle bombardment and *Agrobacterium*-mediated techniques. Plant Cell Rep 23, 780-789.

Van der Biezen, E.A., and Jones, J.D. (1998). Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci 23, 454-456.

Van der Hoorn, R.A., De Wit, P.J., and Joosten, M.H. (2002). Balancing selection favors guarding resistance proteins. Trends Plant Sci 7, 67-71.

Vandepoele, K., Simillion, C., and Van de Peer, Y. (2003). Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15, 2192-2202.

Vicient, C.M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E., and Schulman, A.H. (1999). Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell 11, 1769-1784.

Vision, T.J., Brown, D.G., and Tanksley, S.D. (2000). The origins of genomic duplications in *Arabidopsis*. Science 290, 2114-2117.

Wan, Y., and Lemaux, P.G. (1994). Generation of large numbers of independently transformed fertile barley plants. Plant Physiol 104, 37-48.

Wang, M.B., Abbott, D.C., Upadhyaya, N.M., Jacobsen, J.V., and Waterhouse, P.M. (2001). *Agrobacterium tumefaciens*-mediated transformation of an elite Australian barley cultivar with virus resistance and reporter genes. Aust J Plant Physiol 28, 149-156.

Wang, X., Shi, X., Hao, B., Ge, S., and Luo, J. (2005). Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165, 937-946.

Weibull, J., Walther, U., Sato, K., Habekuss, A., Kopahnke, D., and Proeseler, G. (2003). Diversity in resistance to biotic stress. In Diversity in Barley (*Hordeum vulgare*), R.V. Bothmer, T.V. Hintum, H. Knüpffer, and K. Sato, eds (Amsterdam: Elsevier Science), pp. 143-178.

Wicker, T., Matthews, D.E., and Keller, B. (2002). TREP: A database for Triticeae repetitive elements. Trends Plant Sci 7, 561-562.

Wicker, T., Guyot, R., Yahiaoui, N., and Keller, B. (2003). CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol 132, 52-63.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E., and Keller, B. (2001). Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26, 307-316.

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.D., Dubcovsky, J., and Keller, B. (2003). Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15, 1186-1197.

Xiao, S., Ellwood, S., Calis, O., Patrick, E., Li, T., Coleman, M., and Turner, J.G. (2001). Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. Science 291, 118-120.

Yahiaoui, N., Srichumpa, P., Dudler, R., and Keller, B. (2004). Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37, 528-538.

Yoshimura, S., Yamanouchi, U., Katayose, Y., Toki, S., Wang, Z.X., Kono, I., Kurata, N., Yano, M., Iwata, N., and Sasaki, T. (1998). Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95, 1663-1668.

Zhang, B.H., Pan, X.P., Cobb, G.P., and Anderson, T.A. (2006). Plant microRNA: A small regulatory molecule with big impact. Dev Biol 289, 3-16.

Zhou, F., Kurth, J., Wei, F., Elliott, C., Vale, G., Yahiaoui, N., Keller, B., Somerville, S., Wise, R., and Schulze-Lefert, P. (2001). Cell-autonomous expression of barley *Mla1* confers race-specific resistance to the powdery mildew fungus via a *Rar1*-independent signaling pathway. Plant Cell 13, 337-350.

Zhou, T., Wang, Y., Chen, J.Q., Araki, H., Jing, Z., Jiang, K., Shen, J., and Tian, D. (2004). Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271, 402-415.

Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D., and Cregan, P.B. (2003). Single-nucleotide polymorphisms in soybean. Genetics 163, 1123-1134.

IX. APPENDIX

Appendix 9-1 Detection of H2 sub-haplotypes by PCR

(A) PCR amplification results. 1 indicates an amplification product of the expected size. 0 indicates no amplification. The haplotype column shows the haplotype structure of each line. H1 lines are shown in dark grey. H2 lines that had the same pattern as Renan are displayed in light grey, H2 lines that had the same pattern as Renan for the short PCR products (A, B and C) but not for the large PCR product (B_3kb) are displayed in white, and H2 lines that did not show any amplification apart from the positive control are displayed in middle grey.
(B) Primer pairs used to amplify genomic fragments surrounding *RGA2-a* and *RGA2-b* (cf. Figure 4-6). Control (+) corresponds to a fragment spanning the 3' end of *RGA2*, A and C fragments spanning the disruption and inversion point of *RGA2* and B the deletion point of the beginning of *RGA2*.

A

Lines	Species	Origin	Control (+)	A	C	B	B_3kb	Haplotype
ThLr10	*T. aestivum*	Canada	1	0	0	0	0	H1
DV92	*T. monococcum* L.	Italy	1	0	0	0	0	H1
TRI 17119	*T. urartu*	Turkey	1	0	0	0	0	H1
TRI 656	*T. monococcum* L.	Albania	1	0	0	0	0	H2
TRI 2006	*T. monococcum* L.	Spain	1	1	1	1	1	H2
TRI 4321	*T. monococcum* L.	Marocco	1	0	0	0	0	H2
TRI 13061	*T. monococcum* L.	Italy	1	1	1	1	1	H2
TRI 13605	*T. monococcum* L.	Georgia	1	1	1	1	1	H2
TRI 17434 *	*T. monococcum* L.	Azerbaijan	1	1	1	1	0	H2
TRI 11495	*T. urartu*	Turkey	1	0	0	0	0	H2
TRI 17413	*T. urartu*	Iraq	1	0	0	0	0	H2
TRI 3425 *		Europe	1	1	1	1	1	H2
IG 46516	*T. turgidum* ssp. *dicoccoides*	Syria	1	0	0	0	0	H2
TTD22	*T. turgidum* ssp. *dicoccoides*	Israel	1	1	1	1	1	H2
TTD37	*T. turgidum* ssp. *dicoccoides*	Iran	1	1	1	1	0	H2
TRI 1514 *		Europe	1	1	1	1	1	H2
TRI 2967 *	*T. turgidum* ssp. *dicoccum*	Europe	1	1	1	1	0	H2
TRI 673 *	*T. turgidum* ssp. *durum*	Greece	1	1	1	1	1	H2
TRI 1877 *	*T. turgidum* ssp. *durum*	Greece	1	1	1	1	1	H2
TA 952 *	*T. timopheevii* ssp. *armeniacum*	Iraq	1	1	1	1	1	H2
TA 1008 *	*T. timopheevii* ssp. *armeniacum*	Turkey	1	1	1	1	1	H2
Chinese Spring	*T. aestivum* L.	China	1	1	1	1	1	H2
Frisal	*T. aestivum* L.	Switzerland	1	1	1	1	1	H2
Arina	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Apollo	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Ares	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Altgold	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Basalt	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Bezostaja	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Boval	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Bussard	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Capelle	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Caribo	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Castracht	*T. aestivum* L.	Europe	1	1	1	1	1	H2

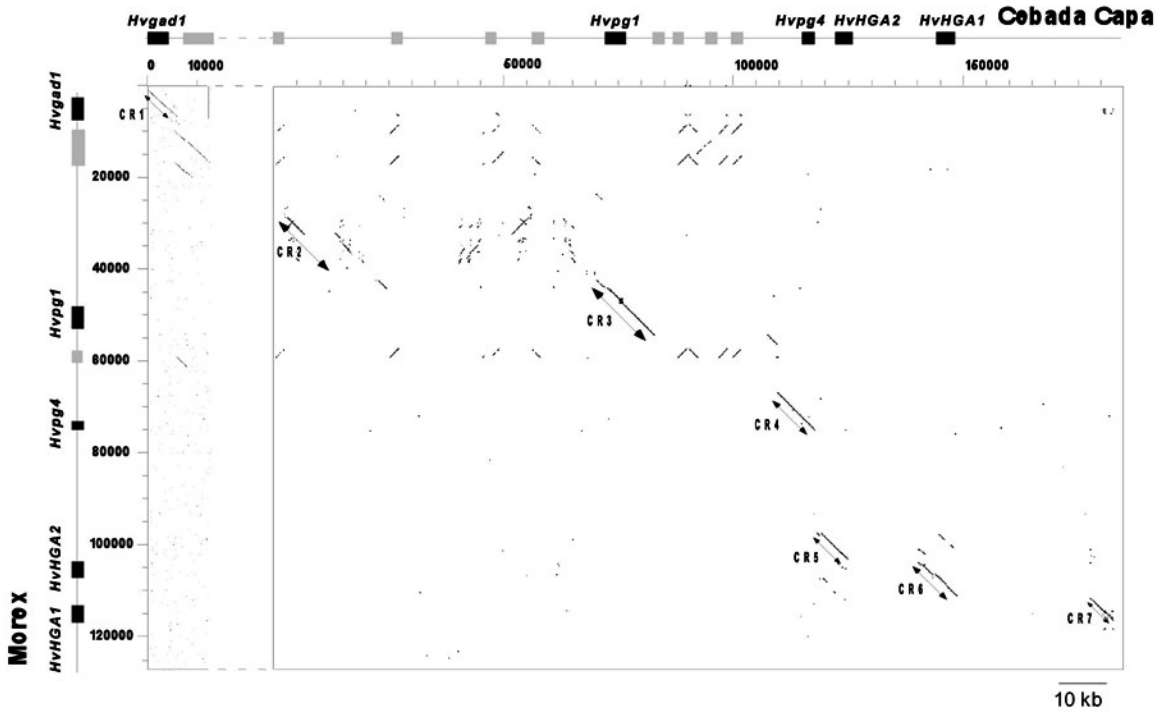
Appendix 9-1 continued

Champlein	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Derenburgersilber	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Disponent	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Galaxie	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Granada	*T. aestivum* L.	Europe	1	1	1	1	1	H2
HeineVII	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Hercule	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Hoeser 52	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Hubel	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Kanzler	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Kavkas	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Komoran	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Kraka	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Kronjuwel	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Lona	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Lueg	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Maris Huntsmen	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Merlin	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Monopol	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Obelisk	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Oberkulmer	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Osmut	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Ostra	*T. aestivum* L.	Europe	1	1	1	1	1	H2
Probus	*T. aestivum* L.	Europe	1	1	1	1	0	H2
Renan	*T. aestivum* L.	Europe	1	1	1	1	1	H2

*: Lines from which the B fragment was sequenced.

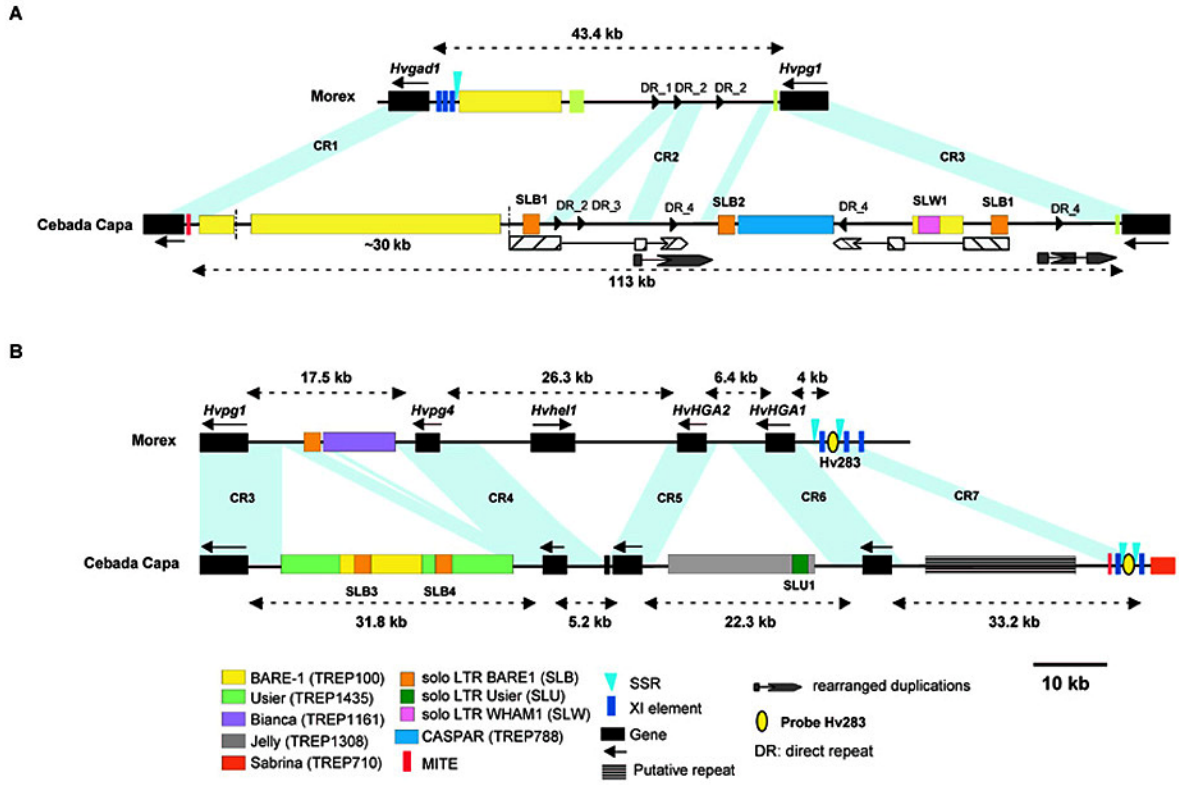
B

Primers	Forward	Reverse	Expected product size	Annealing temperature
Control (+)	5'-ACACATGTTCCATCCAACGG	5'-CTGGATATCCTCGTGAGCAT	477 bp	52°C
A	5'-AGCTGCAACCTTCCTCCAAT	5'-GCTTATAGATTCGCCTCCCAA	415 bp	53°C
C	5'-AAGCTCAAACGTTTGTTGCGG	5'-GCTAAAAGGTTGATGTCGGAC	474 bp	52°C
B	5'-ACAAGACCCCAGGATAGAGG	5'-GTGCGTCATTGAGTTCCAGA	331 bp	52°C
B_3kb	5'-GTCTCCAAGGCCACATTGAA	5'-GTGCGTCATTGAGTTCCAGA	3 kb	52°C



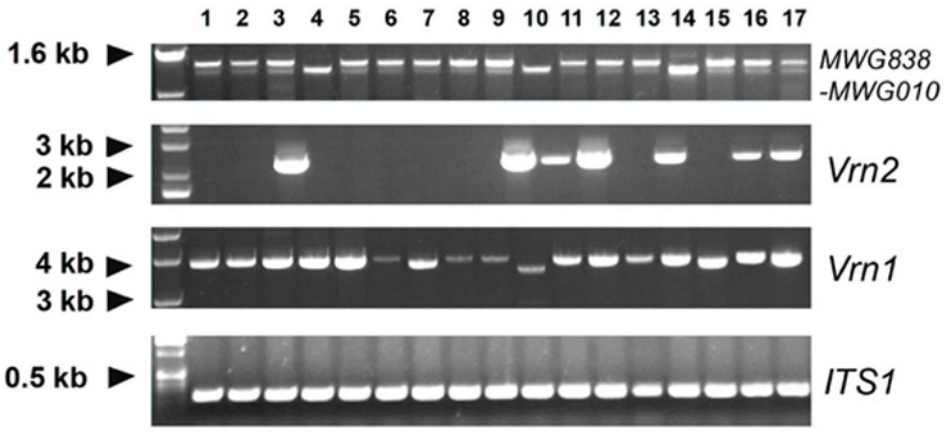
Appendix 9-2 Dot plot comparison of the orthologous sequences from Morex and Cebada Capa

126.6 kb of sequence from Morex (AF521177, position 85,000 to 211,664) was aligned with 12,258 bp (AY642925) and 184,426 bp (AY642926) from Cebada Capa. The lines above each axis represent the entire region in each cultivar. Genes are indicated as black boxes and BARE-1 LTR retrotransposons are in grey. Seven conserved regions (CR) are indicated as lines delimited by arrowheads. The interrupted line represents the ~30 kb of unassembled BARE-1 sequence in Cebada Capa.



Appendix 9-3 Schematic representation of the comparison between the orthologous *Hvgad1-Hvpg1* intergenic regions (A) and the *Hvpg1 and Hv283* interval (B) in Morex and Cebada Capa

The seven conserved regions (CR) are indicated with light blue areas. The size of the intergenic regions in each of the two sequences is indicated between double arrowheads. The partially sequenced 30 kb region containing several BARE-1 elements is shown between two vertical dotted lines. The rearranged duplications of 21 kb and 9 kb in the CR2-CR3 interval in Cebada Capa are shown below the annotation. Genes are indicated as black boxes with their name above them and their transcriptional orientation is shown with an arrow. Repetitive elements are drawn as boxes with the color code indicated in the legend. Complete sequence annotations are given under the accession numbers AF521177 for Morex and AF642925-26 for Cebada Capa.



Appendix 9-4 PCR amplification on 17 barley cultivars representative for the different haplotypes found at three independent loci

Line 1: Bowman, 2: Bowman5*/Clipper BC67, 3: Triumph, 4: Bowman*/PI531849, 5: Egypt, 6: Sundance, 7: Tunisian, 8: Elisa, 9: Ellinor, 10: Franka, 11: Hanka, 12: Hor4445, 13: JCJ-188, 14: L94, 15: Quinn, 16: Michka, 17: Pallas.

Appendix 9-5 List of primers used for the amplification of intergenic regions at the *Rph7, Vrn1, Vrn2,* and MWG838-MWG010 loci on 41 cultivated barley lines

Locus	Primers	Sequence (5'-3')	T_m [(a)]
Rph7	CR3-1	GGTCCAGACAGTATGTTGAGATAATAAG	58°C
Rph7	CR4-1	GGTTGTGAGATTTGAATTGAGC	54°C
Rph7	CR4-3	CAGGAGCTGGGCAGTAACC	60°C
Rph7	CR5-2	GCAGGGATGACGACAGAGAT	58°C
Rph7	CR4-2	ACCTTCTTCGAGGTCGCCAT	58°C
Rph7	CR5-1	CCACGTGTGAGTACCTAGTTGTCA	60°C
Vrn1	635P2_G4F	GCCGAAGTAGAACCCAGCAG	60°C
Vrn1	635P2_G5R	ATGCGCAGCTCCCTCACCAT	60°C
Vrn2	615K1_G2F	TGAACTCTAAACCTCCGCAT	60°C
Vrn2	615k1_G3R	GCGCAGCAGTTTGAGGAGA	57°C
MWG838-MWG010	519G04_C1F	CGCCAAGAACCGCTACAC	57°C
MWG838-MWG010	519G04_C2R	TGAATGATGGCCTGTGGAA	53°C

(a) Annealing temperature

X. ACKNOWLEDGEMENTS

I wish to thank…

Prof. Dr. Beat Keller who gave me the opportunity to do this PhD thesis in his research group for his support, advises and the confidence he placed in me during these years.

Dr. Catherine Feuillet, my supervisor nearby and from a distance. She always supported and encouraged me and found the right words to cheer me up when times were difficult. Without her, this work would not have succeeded.

Prof. Dr. Bruce McDonald for being member of my PhD committee.

Kathrin Michel who became my best friend during the last years. Kathrin, it was great to have someone at my side who experienced the same highs and lows which a PhD thesis bring along. I am looking forward to many more years of friendship.

Nabila Yahiaoui for her advises and support when things in the lab didn't work as expected.

Barbara Schellenberg and Daniel Stutz for their friendship and the many relaxing evenings at the movies.

Navreet Kaur, Silvia Travella, Susanne Brunner and all other present and former members of the Keller and Dudler group for the nice atmosphere in the lab. I am especially thankful to Röbi Dudler for the encouragement during difficult periods and for all the discussions about any aspect of life.

Theres Imhof-Klimm, Urs Ochsner, Bea Senger, and Geri Herren for their good hands in the lab and the greenhouse.

All my friends who supported me during this time.

Last but no least danke ich meinen Eltern und meinen Geschwistern Judith, Silvia und Adrian. Sie haben mich während all den Jahren meines Studiums ermutigt und unterstützt und waren in jeder Situation für mich da.

XI. CURRICULUM VITAE

Name	Beatrice Scherrer
Date of birth	May 18, 1973
Place of origin	Willisau LU
1980 – 1986	Primary school in Willisau
1986 – 1989	Secondary school in Willisau
1989 – 1992	Commercial apprenticeship with the municipal administration in Willisau
1993 – 1997	Grammar school for adults (MSE) in Luzern; Matura Type B (Latin)
1997 – 2002	Undergraduate studies in biology at the University of Zürich
April 2002	Diploma in plant biology (major) and molecular biology (minor)
2002 – 2006	PhD thesis at the Institute of Plant Biology at the University of Zürich, Group of Prof. Beat Keller